nature

DEEP VIRAL IMPACT

Virus infections in sea-bed microbes feed the world's largest ecosystem

THE HUMAN LIFESPAN
How far can you go?

MILKY WAY SATELLITES
Galaxies with a dark side

ONCOGENESIS
The natural selection of cancers

nature
www.nature.com/nature
Vol 454 | Issue no. 7208 | 28 August 2008

Doubly endangered

The landmark Endangered Species Act in the United States needs more flexibility and fresh thinking — but not of the kind being advocated by the Bush administration.

In 1995, then-US Senator Dirk Kempthorne (Republican, Idaho) wrote to *The New York Times* decrying accusations that his effort to reform the Endangered Species Act was actually an attempt to dismantle it. The debate had become "so polarized", Kempthorne wrote, that anybody advocating change would be "quickly tagged as anti-environment".

He had a point. The act, among the most wide-ranging environmental laws in the world, is designed to protect any critically endangered species in the United States from extinction by any form of economic development. It was passed in 1973, when environmental idealism pretty much trumped every other consideration — not a single senator voted nay. But opposition built quickly, as people began to realize that the word 'species' applied not just to photogenic bald eagles, but to fish, rodents, plants, insects and a host of other organisms they had never heard of. The snail darter fish and the spotted owl could stop major dam-building and logging operations in their tracks. Many Americans came to see the act as a symbol of federal regulation run amok — even as committed environmentalists came to see it as the ultimate backstop, the statute whose powers could be invoked when all other planning and conservation tools failed.

Thus Kempthorne's frustration. His 1995 effort in the Senate to reform the act was ultimately futile, but it was a serious attempt to seek common ground on the issue.

Thirteen years later, Kempthorne is Secretary of the Interior in the Bush administration and he can no longer make that claim. His latest efforts to alter the act are neither benign nor bipartisan. Currently, federal agencies giving permission for any project such as a road, a dam or logging must consult with the US Fish and Wildlife Service to ensure that no endangered species are at risk. But under Kempthorne's latest proposal, announced earlier this month, agencies would be allowed to skip that step if their own analysis finds such action unnecessary. In the wrong hands, this authority would effectively end oversight on projects and promote the relentless development that is steadily consuming the world's remaining natural habitats.

Exaggerated burden

There is a lot to be said for addressing the Fish and Wildlife Service's workload; its biologists are swamped with tens of thousands of project consultations each year. But that problem is best addressed by giving the agency more money and people, and perhaps by finding ways for other agencies to share the paperwork — not by gutting the concept of scientific review.

More broadly, critics of the act should concede that its burdens have been greatly exaggerated. Although the law has a powerful regulatory hammer, it is seldom deployed; permits for projects known to harm species or their habitat are granted all the time. Nor is the act inflexible. Administrators have developed creative mechanisms that provide incentives for landowners to manage their land to benefit threatened or endangered species.

At the same time, environmentalists who tirelessly defend the act would be wise to acknowledge its faults. It has indeed helped to rescue high-profile species from the brink of extinction, and served as the catalyst for conservation efforts that go well beyond individual species. One example is regulation of the Edwards Aquifer in semi-arid southern Texas, the water source not only for the fountain darter fish and other imperilled species but also for 1.7 million Texans. But the numerous controversies engendered by the act have sparked a backlash that has set back the cause of conservation. "Shoot, shovel and shut up," goes a common rural refrain.

> "Environmentalists who tirelessly defend the act would be wise to acknowledge its faults."

Moreover, protection under the act for individual species and their habitat is generally doled out by judges, not biologists. And when it comes to habitat loss in general — the biggest threat — the act is quite constrained. Although protecting 'indicator' or 'keystone' species can bring about broader habitat protection that benefits other species, this the approach is still, at its core, purely reactive.

Looking forward

What is needed is a proactive system for evaluating and protecting habitat and biodiversity and the benefits they bring to people. How to go about that task is a difficult question, but one that the international biodiversity community is actively pursuing. One proposal is for a biodiversity science panel under the United Nations, similar to that for global warming. Another is a biodiversity monitoring network to gather hard data on a global scale. We commend and encourage these efforts. Nevertheless, each nation will ultimately need to apply any lessons learned within their own legal frameworks.

Were policy-makers in the United States to start from scratch and rewrite the Endangered Species Act today, they would probably come up with something quite different — including an emphasis on the kind of broad habitat plans that would protect healthy species before they begin to decline.

Unfortunately, amending the act has proved almost impossible (the last significant change was in 1982). But even within the existing framework, there is room for the next administration to manoeuvre. Former interior secretary Bruce Babbitt, who served under President Bill Clinton, initiated the first conservation plans designed to protect multiple species — including many that were not yet listed. And although subsequent implementation and monitoring have fallen short on some of these plans, the principle remains sound.

In this context, the fact that creative regulations have been implemented under the existing statute might well be the only cause for optimism in an otherwise stale debate. ■

After Musharraf

Pakistan's elected governments should break the habit of a lifetime and give due priority to science.

Military strongmen have ruled Pakistan for longer than elected politicians — and, paradoxically, have treated that nation's scientists far better than the much less stable civilian administrations.

General Pervez Musharraf, whose nine-year rule ended with his resignation on 18 August, was a prime example: his regime greatly strengthened the foundations for a Pakistani knowledge economy, instituting reforms that included bigger research budgets, an ambitious university-building programme, a nationwide digital library, a scheme to attract international faculty, and performance-related pay for professors. Many of the changes have been praised in external evaluations within the past year, including those of the World Bank, the US Agency for International Development and the British Council.

But a week after Musharraf was forced to go, the governing coalition of political parties that replaced him has collapsed. Worryingly, Pakistan's governance of its science seems all set to revert to the situation that prevailed under previous elected governments, when the science and education ministries were often afterthoughts — and, in the case of science, with laughable budgets, no ministers and no leadership.

Is history repeating itself? Pakistan has lacked both a science and an education minister for several months, important projects are on hold and budgets have been cut. Only one out of nine planned new universities with international partners is to go ahead on schedule and, in spite of earlier promises to the contrary, the higher-education budget for the fiscal year 2007–08 has been cut from US$438 million to US$364 million. Vice-chancellors at 12 universities have had to take out bank loans to pay salaries.

The future of Atta-ur-Rahman, the former president's long-serving science adviser, is also uncertain. Atta-ur-Rahman, a professor of natural product chemistry at the University of Karachi, was the chief architect of reforms under Musharraf, who once boasted that he had so much confidence in his adviser that Atta-ur-Rahman's ideas would be funded "without reservation".

But amid the turmoil, there are reasons for optimism. Nina Fedoroff, science adviser to US Secretary of State Condoleezza Rice, was in Islamabad earlier this week for talks on ramping up US assistance to Pakistan's science programmes. The United States and Pakistan together contribute $150 million to what is the world's largest Fulbright programme of fellowships, enabling Pakistani students to go to the United States for postgraduate study and research. In addition, Pakistan is slated to receive a further $7.5 billion in US aid over the next five years and Pakistani officials would like 20% of this earmarked for science and innovation.

Another reason to be optimistic is that this round of non-military government could be stabler and stronger than previous ones. The incoming government will be dominated by the Pakistan Peoples Party, led by Asif Ali Zardari, the widower of the murdered ex-Prime Minister Benazir Bhutto. Should Zardari become the new president, this will bring a measure of stability and an opportunity for real change. Previous elected governments dropped the ball on science and education because they were weak and spent a lot of effort — unsuccessfully — keeping the generals at bay. Zardari's party has a majority in parliament and Pakistan's military is so far showing little interest in returning to politics.

But the Peoples Party has some learning to do. Its previous record on science is among the most misguided of all Pakistan's elected governments, and Zardari himself has served time in prison on corruption charges — which he denies. In the late 1970s, the party's founder Zulfikar Ali Bhutto diverted scarce resources and personnel into building the nuclear bomb. His daughter Benazir's contribution during the 1980s and 1990s was a science ministry without a dedicated minister.

A return to the pre-Musharraf era would send Pakistan back to the scientific stone age. The new government needs to recognize that regardless of how much it disliked him, the general bequeathed it a foundation in science and technology on which it can build. ■

Future transport

The hike in the price of oil means that new ways of fuelling transport are no longer fantasy.

Today's globalized economy largely rests on nineteenth- and twentieth-century revolutions in transport: humankind's ability to move goods and people around the planet by boat, train, car and plane. The global transportation network allows consumers to buy crisp New Zealand apples in London, fresh seafood in Oklahoma City and Chinese-manufactured goods everywhere.

Indeed, transportation is so integral to the global economy that 14% of the world's greenhouse-gas emissions come from that sector alone. And in the developing countries especially, that fraction is growing rapidly. Witness the explosion of private transport, which has always been a symbol of wealth: Beijing gets another 1,000 cars every day. Earlier this year, India's Tata Motors unveiled the world's cheapest car, a $2,500 bubble called the Nano. There is every reason to expect that car ownership will continue to increase towards that of the already developed nations, where between a third and half of the population owns a car.

This is why managing the greenhouse emissions from transportation has emerged as a major challenge in the twenty-first century. In a series of Special Reports (see http://tiny.cc/CyqtP) that concludes this week, *Nature* has looked at technologies that could help accomplish that goal. Some are relatively familiar, such as fuel cells running on hydrogen split from water via solar or wind power. Others have a definite back-to-the-future feel, such as kite-powered ships or steam-powered trains (see page 1036). All could have a significant impact.

Not so long ago most of these ideas would have been dismissed as pipe dreams. Today, if the soaring price of oil has a silver lining, it is that the push for alternative transportation technologies has become real and serious. Major research investments are being made by government, industry and venture capitalists. But oil prices are notoriously volatile; at this particular moment they seem to be declining. The challenge for policy-makers in every sector is to make sure those investments in future transportation are sustained for the long haul. ■

RESEARCH HIGHLIGHTS

ELECTRONICS

Silicon enhancement

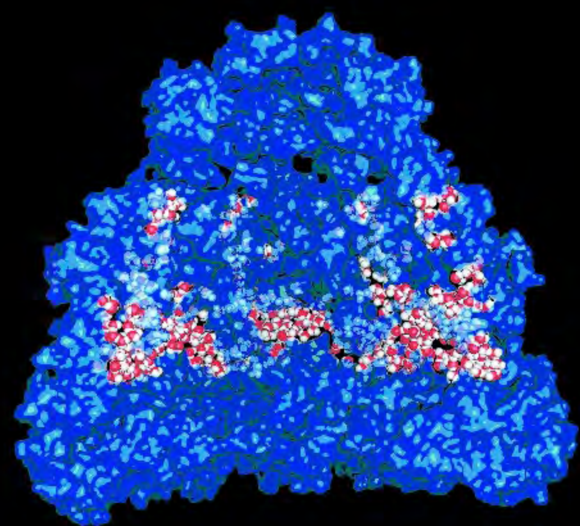
Science **321,** 1069–1071 (2008)

The formation of compounds containing silicon usually involves the silicon atoms changing their oxidation states. But Gregory Robinson, Paul von Schleyer and their colleagues at the University of Georgia in Athens have produced a stable silicon compound in which the silicon stays in its 'zero oxidation state'. Silicon atoms in this state are normally highly reactive, but on this occasion are attached to large carbon-based rings that stabilize them.

The researchers verified this outcome with X-ray crystallography: if the silicon atoms were in an oxidation state of '+2', the molecule would be flat, but it is bumpy. And from bond-length measurements, it seems that the two silicon atoms in the compound are connected to each other by a double bond. They have one free pair of electrons and another pair donated by the organic attachment.

Water bomb

J. Am. Chem. Soc. doi:10.1021/ja802248m (2008)



J. L. ENGLAND ET AL./J. AM. CHEM. SOC.

Water trapped inside barrel-shaped enzymes called chaperonins could be crucial to the way they help proteins fold. Without such attendance, complicated proteins would fail to form their proper arrangement, and consequently would not work.

Vijay Pande and his colleagues at Stanford University in California followed the change in shape that exposes hydrophilic chemical groups on the inner surface of a chaperonin complex called GroE (pictured) as the enzyme takes in an unfolded protein. Using computer simulations, they suggest this process enables GroE to sequester minute amounts of water (red and white) in its active site. This water creates a driving force for the protein to fold, burying its hydrophobic surfaces in the interior of the structure that it eventually forms.

EVOLUTIONARY BIOLOGY

Commonality and cuckoos

Behav. Ecol. Sociobiol. doi:10.1007/s00265-008-0618-0 (2008)

Natural selection is driving birds that are parasitized by common cuckoos (*Cuculus canorus*) to lay clutches of more uniformly patterned eggs, researchers have found.

Working around the village of Apaj near Kiskunság National Park in Hungary, Csaba Moskát of the Hungarian Academy of Sciences and his co-workers painted different numbers of specks onto the first three eggs laid in ten great reed warbler (*Acrocephalus arundinaceus*, pictured below) nests. They then painted the fourth eggs to arrive in those nests and in 21 others to look like parasitic eggs.

M. VARESVUO/NATUREPL.COM

Egg uniformity foils cuckoos, the warblers' behaviour revealed. Warblers with eggs manipulated to be more different from one another tolerated the fake parasitic egg 40% of the time; that figure was just 5% for the 21 control nests.

MICROBIOLOGY

Suffocating tuberculosis

Proc. Natl Acad. Sci. USA **105,** 11945–11950 (2008)

A compound that inhibits the production of ATP, the primary energy carrier in cells, could make treating tuberculosis a little easier, report Kevin Pethe at the Novartis Institute for Tropical Diseases in Chromos, Singapore, and his colleagues. The disease-causing bacterium *Mycobacterium tuberculosis* can evade treatment by entering into a quiescent, non-dividing state that is resistant to current therapies.

Pethe and his team found that ATP levels in these bacteria are five- to sixfold lower than normal, but that quiescent *M. tuberculosis* does still require ATP to survive. They reasoned that this could render the microbe particularly susceptible to drugs that inhibit ATP synthesis. And, as it turns out, one such inhibitor that acts in a dose-dependent manner, R207910, kills quiescent *M. tuberculosis* at levels that don't kill other cells.

PARTICLE PHYSICS

Antimatter bounces back

Phys. Rev. A **78,** 022506 (2008)

In the early 1990s, physicists at CERN in Switzerland watched as antiprotons and helium annihilated in flashes of energy. But their experiment also yielded an unexplained secondary string of annihilations in the facility's experimental chamber.

Andrea Bianconi of Italy's National Institute of Nuclear Physics in Brescia and his colleagues have now modelled what happened. Their calculations show that some antiprotons bounced off the aluminium back wall before striking helium atoms. Their model, which matches the data, shows that roughly a quarter of the antiprotons were reflected by aluminium nuclei. Not all matter–antimatter interactions end with a bang, it seems.

EVOLUTION

Serotonin for mothers

Nature Neurosci. doi:10.1038/nn.2176 (2008)

The neurotransmitter serotonin is known to be important in mood and behaviour; now researchers have shown that its function is also essential to the survival of baby mice.

Evan Deneris at Case Western Reserve University in Cleveland, Ohio, and his colleagues compared the reproductive success of normal mice with that of animals

RESEARCH HIGHLIGHTS

ELECTRONICS

Silicon enhancement

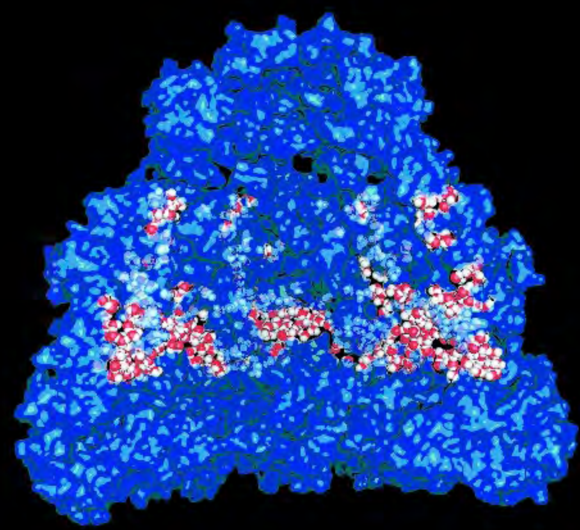
Science **321,** 1069–1071 (2008)

The formation of compounds containing silicon usually involves the silicon atoms changing their oxidation states. But Gregory Robinson, Paul von Schleyer and their colleagues at the University of Georgia in Athens have produced a stable silicon compound in which the silicon stays in its 'zero oxidation state'. Silicon atoms in this state are normally highly reactive, but on this occasion are attached to large carbon-based rings that stabilize them.

The researchers verified this outcome with X-ray crystallography: if the silicon atoms were in an oxidation state of '+2', the molecule would be flat, but it is bumpy. And from bond-length measurements, it seems that the two silicon atoms in the compound are connected to each other by a double bond. They have one free pair of electrons and another pair donated by the organic attachment.

Water bomb

J. Am. Chem. Soc. doi:10.1021/ja802248m (2008)



J. L. ENGLAND ET AL./J. AM. CHEM. SOC.

Water trapped inside barrel-shaped enzymes called chaperonins could be crucial to the way they help proteins fold. Without such attendance, complicated proteins would fail to form their proper arrangement, and consequently would not work.

Vijay Pande and his colleagues at Stanford University in California followed the change in shape that exposes hydrophilic chemical groups on the inner surface of a chaperonin complex called GroE (pictured) as the enzyme takes in an unfolded protein. Using computer simulations, they suggest this process enables GroE to sequester minute amounts of water (red and white) in its active site. This water creates a driving force for the protein to fold, burying its hydrophobic surfaces in the interior of the structure that it eventually forms.

EVOLUTIONARY BIOLOGY

Commonality and cuckoos

Behav. Ecol. Sociobiol. doi:10.1007/s00265-008-0618-0 (2008)

Natural selection is driving birds that are parasitized by common cuckoos (*Cuculus canorus*) to lay clutches of more uniformly patterned eggs, researchers have found.

Working around the village of Apaj near Kiskunság National Park in Hungary, Csaba Moskát of the Hungarian Academy of Sciences and his co-workers painted different numbers of specks onto the first three eggs laid in ten great reed warbler (*Acrocephalus arundinaceus*, pictured below) nests. They then painted the fourth eggs to arrive in those nests and in 21 others to look like parasitic eggs.

M. VARESVUO/NATUREPL.COM

Egg uniformity foils cuckoos, the warblers' behaviour revealed. Warblers with eggs manipulated to be more different from one another tolerated the fake parasitic egg 40% of the time; that figure was just 5% for the 21 control nests.

MICROBIOLOGY

Suffocating tuberculosis

Proc. Natl Acad. Sci. USA **105,** 11945–11950 (2008)

A compound that inhibits the production of ATP, the primary energy carrier in cells, could make treating tuberculosis a little easier, report Kevin Pethe at the Novartis Institute for Tropical Diseases in Chromos, Singapore, and his colleagues. The disease-causing bacterium *Mycobacterium tuberculosis* can evade treatment by entering into a quiescent, non-dividing state that is resistant to current therapies.

Pethe and his team found that ATP levels in these bacteria are five- to sixfold lower than normal, but that quiescent *M. tuberculosis* does still require ATP to survive. They reasoned that this could render the microbe particularly susceptible to drugs that inhibit ATP synthesis. And, as it turns out, one such inhibitor that acts in a dose-dependent manner, R207910, kills quiescent *M. tuberculosis* at levels that don't kill other cells.

PARTICLE PHYSICS

Antimatter bounces back

Phys. Rev. A **78,** 022506 (2008)

In the early 1990s, physicists at CERN in Switzerland watched as antiprotons and helium annihilated in flashes of energy. But their experiment also yielded an unexplained secondary string of annihilations in the facility's experimental chamber.

Andrea Bianconi of Italy's National Institute of Nuclear Physics in Brescia and his colleagues have now modelled what happened. Their calculations show that some antiprotons bounced off the aluminium back wall before striking helium atoms. Their model, which matches the data, shows that roughly a quarter of the antiprotons were reflected by aluminium nuclei. Not all matter–antimatter interactions end with a bang, it seems.

EVOLUTION

Serotonin for mothers

Nature Neurosci. doi:10.1038/nn.2176 (2008)

The neurotransmitter serotonin is known to be important in mood and behaviour; now researchers have shown that its function is also essential to the survival of baby mice.

Evan Deneris at Case Western Reserve University in Cleveland, Ohio, and his colleagues compared the reproductive success of normal mice with that of animals

lacking many of the neurons that produce serotonin. Litters in the care of mothers with low serotonin levels died within a few days of birth despite adequate nursing.

The mothers' behaviour explained this: they built poor-quality nests and did not keep their offspring huddled together, leaving the litter exposed to the cold. When these mothers' young were fostered by normal mothers immediately after birth, their odds of living rose to normal.

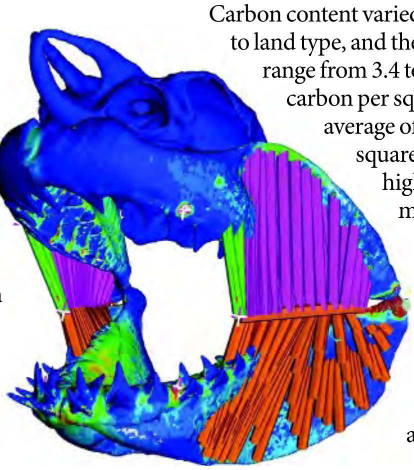
PALAEOBIOLOGY

Megabite

J. Zool. doi:10.1111/j.1469-7998.2008.00494.x (2008)

How hard can a great white bite? The shark (*Carcharodon carcharias*) can chomp with about 18,000 Newtons of downward force — the strongest known bite of any living species — and its habit of shaking its head from side to side might increase this further.

Stephen Wroe at the University of New South Wales in Sydney, Australia, and his colleagues calculated this flesh-ripping force from a three-dimensional computer model of the animal's skull, jaw and muscles (pictured right). Sharks' skeletons are made of cartilage rather than the much stiffer bone, and although this causes the jaws to undergo greater deformation, it seems to be no impediment to a powerful bite.



S. WROE ET AL.

The team estimates that the even larger extinct shark *C. megalodon*, which is thought to have hunted whales, may have bitten with an order of magnitude more force.

MATERIALS SCIENCE

Finding focus

Appl. Phys. Lett. **93,** 05311 (2008)

Physicists would like to make lenses from metamaterials — structures that can have a negative refractive index. Such lenses would tightly focus light over short distances with little or no distortion. So far, researchers have only been able to make crude versions of these lenses from aluminium oxide rods. But Didier Casse, Srinivas Sridhar and their colleagues at Northeastern University in Boston, Massachusetts, say they have built a two-dimensional metamaterials lens from a semiconductor crystal.

The lens was made by carving a lattice into a layered indium gallium arsenic phosphide wafer. It can focus infrared light of the same wavelength as that used in common devices to a spot just 12 micrometres away from its surface. The authors say that the work could lead to smaller, more efficient digital cameras and sensors.

GEOSCIENCES

Soil sink surprise

Nature Geosci. doi:10.1038/ngeo284 (2008)

It is vital to know how much carbon is stored in Earth's Arctic soils, because much of it could be released as greenhouse gases as the planet warms.

Chien-Lu Ping at the University of Alaska Fairbanks in Palmer and his colleagues dug 117 one-metre-deep pits across the North American Arctic region to gather data on this pressing problem. They combined this with 22 existing datasets from shallower holes.

Carbon content varied markedly according to land type, and the team's measurements range from 3.4 to 55.1 kilograms of carbon per square metre. Their average of 34.8 kilograms per square metre is substantially higher than previous measurements of between 20 and 29 kilograms per square metre, and suggests that current estimates of the total carbon stored in Arctic soils are too low.

ZOOLOGY

Under pressure

Curr. Biol. **18,** R695–R696 (2008)

Wind turbines are bafflingly bad for bats, the bodies of which are often found beneath their blades. Erin Baerwald and her colleagues at the University of Calgary in Canada think they know why: the pressure differential near the blades causes decompression sickness in the animals.

The team collected dead hoary bats (*Lasiurus cinereus*) and silver-haired bats (*Lasionycteris noctivagans*) at a wind facility in Alberta. During autopsy, they found that 69 of the 75 bats had suffered haemorrhaging in the chest or abdomen, even though 32 seemed to have no external injuries. The remaining six had an external injury but no evidence of internal bleeding.

Bats apparently encounter the lethal pressure change as they flit close to the turbine blades.

JOURNAL CLUB

John Harte
University of California, Berkeley

An ecologist notes that important details are missing from climate-change models.

Unmitigated climate change will gravely reduce Earth's biodiversity. How much this will happen is calculated by combining data on how the species richness of different habitats varies with their area and projections of how much various habitat types will shrink as the planet warms.

But such grand analyses are blunt instruments; they miss numerous local processes. I have seen, for example, rosy finches and ptarmigans feeding on the contracting ring of vegetation that surrounds melting snow patches on Alpine slopes. Would these creatures survive the summer if the snow patches melted in late spring rather than late summer? Formed in existing mountainside hollows, snowbeds will not march uphill as the climate warms.

This question was recently answered by Robert Björk and Ulf Molau, then both at the University of Gothenburg in Sweden. They reviewed how the release of water and nutrients from the contracting edge of lingering snow patches sustains alpine life in midsummer by providing nourishing vegetation (R. Bjork and U. Molau *Arctic Antarctic Alpine Res.* **39,** 34–43; 2007). The duo propose that bryophytes, grasses, sedges and rushes will be worst hit by the patches' earlier annual disappearance, and that these easy-to-graze species will be replaced by shrubs and trees, hitting Alpine herbivores hard.

This is just one example of the many impacts on biodiversity that fall through the cracks of current, coarse projections. Life and climate intersect on fine spatial and temporal scales — in the microclimates provided by terrestrial 'nurse plants' and in rock pools that form fleetingly in bedrock depressions. The disruption of these delicate intersections may add up to even more damage to biodiversity than the large-scale models predict. This deserves more study.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

lacking many of the neurons that produce serotonin. Litters in the care of mothers with low serotonin levels died within a few days of birth despite adequate nursing.

The mothers' behaviour explained this: they built poor-quality nests and did not keep their offspring huddled together, leaving the litter exposed to the cold. When these mothers' young were fostered by normal mothers immediately after birth, their odds of living rose to normal.

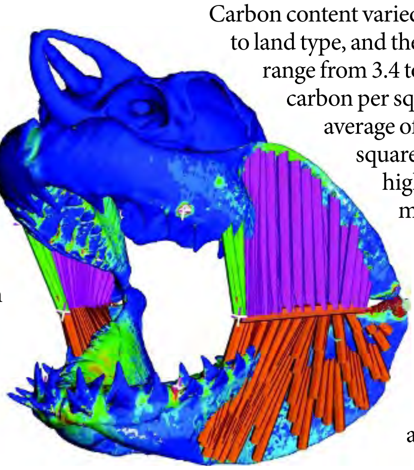
PALAEOBIOLOGY

Megabite

J. Zool. doi:10.1111/j.1469-7998.2008.00494.x (2008)

How hard can a great white bite? The shark (*Carcharodon carcharias*) can chomp with about 18,000 Newtons of downward force — the strongest known bite of any living species — and its habit of shaking its head from side to side might increase this further.

Stephen Wroe at the University of New South Wales in Sydney, Australia, and his colleagues calculated this flesh-ripping force from a three-dimensional computer model of the animal's skull, jaw and muscles (pictured right). Sharks' skeletons are made of cartilage rather than the much stiffer bone, and although this causes the jaws to undergo greater deformation, it seems to be no impediment to a powerful bite.



S. WROE ET AL.

The team estimates that the even larger extinct shark *C. megalodon*, which is thought to have hunted whales, may have bitten with an order of magnitude more force.

MATERIALS SCIENCE

Finding focus

Appl. Phys. Lett. **93,** 05311 (2008)

Physicists would like to make lenses from metamaterials — structures that can have a negative refractive index. Such lenses would tightly focus light over short distances with little or no distortion. So far, researchers have only been able to make crude versions of these lenses from aluminium oxide rods. But Didier Casse, Srinivas Sridhar and their colleagues at Northeastern University in Boston, Massachusetts, say they have built a two-dimensional metamaterials lens from a semiconductor crystal.

The lens was made by carving a lattice into a layered indium gallium arsenic phosphide wafer. It can focus infrared light of the same wavelength as that used in common devices to a spot just 12 micrometres away from its surface. The authors say that the work could lead to smaller, more efficient digital cameras and sensors.

GEOSCIENCES

Soil sink surprise

Nature Geosci. doi:10.1038/ngeo284 (2008)

It is vital to know how much carbon is stored in Earth's Arctic soils, because much of it could be released as greenhouse gases as the planet warms.

Chien-Lu Ping at the University of Alaska Fairbanks in Palmer and his colleagues dug 117 one-metre-deep pits across the North American Arctic region to gather data on this pressing problem. They combined this with 22 existing datasets from shallower holes.

Carbon content varied markedly according to land type, and the team's measurements range from 3.4 to 55.1 kilograms of carbon per square metre. Their average of 34.8 kilograms per square metre is substantially higher than previous measurements of between 20 and 29 kilograms per square metre, and suggests that current estimates of the total carbon stored in Arctic soils are too low.

ZOOLOGY

Under pressure

Curr. Biol. **18,** R695–R696 (2008)

Wind turbines are bafflingly bad for bats, the bodies of which are often found beneath their blades. Erin Baerwald and her colleagues at the University of Calgary in Canada think they know why: the pressure differential near the blades causes decompression sickness in the animals.

The team collected dead hoary bats (*Lasiurus cinereus*) and silver-haired bats (*Lasionycteris noctivagans*) at a wind facility in Alberta. During autopsy, they found that 69 of the 75 bats had suffered haemorrhaging in the chest or abdomen, even though 32 seemed to have no external injuries. The remaining six had an external injury but no evidence of internal bleeding.

Bats apparently encounter the lethal pressure change as they flit close to the turbine blades.

JOURNAL CLUB

John Harte
University of California, Berkeley

An ecologist notes that important details are missing from climate-change models.

Unmitigated climate change will gravely reduce Earth's biodiversity. How much this will happen is calculated by combining data on how the species richness of different habitats varies with their area and projections of how much various habitat types will shrink as the planet warms.

But such grand analyses are blunt instruments; they miss numerous local processes. I have seen, for example, rosy finches and ptarmigans feeding on the contracting ring of vegetation that surrounds melting snow patches on Alpine slopes. Would these creatures survive the summer if the snow patches melted in late spring rather than late summer? Formed in existing mountainside hollows, snowbeds will not march uphill as the climate warms.

This question was recently answered by Robert Björk and Ulf Molau, then both at the University of Gothenburg in Sweden. They reviewed how the release of water and nutrients from the contracting edge of lingering snow patches sustains alpine life in midsummer by providing nourishing vegetation (R. Bjork and U. Molau *Arctic Antarctic Alpine Res.* **39,** 34–43; 2007). The duo propose that bryophytes, grasses, sedges and rushes will be worst hit by the patches' earlier annual disappearance, and that these easy-to-graze species will be replaced by shrubs and trees, hitting Alpine herbivores hard.

This is just one example of the many impacts on biodiversity that fall through the cracks of current, coarse projections. Life and climate intersect on fine spatial and temporal scales — in the microclimates provided by terrestrial 'nurse plants' and in rock pools that form fleetingly in bedrock depressions. The disruption of these delicate intersections may add up to even more damage to biodiversity than the large-scale models predict. This deserves more study.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

NEWS

U. SINAI/GETTY IMAGES

Tanks roll on the streets of Gori.

Georgian science pays price of conflict

The brief war in Georgia has paralysed research throughout the country, halting a promising resurgence in science there.

Seventy-two research projects have been stopped as a result of the Russian–Georgian conflict that erupted this month, says Natia Jokhadze, director of the Georgia National Science Foundation (GNSF) in Tbilisi. That represents 30% of the foundation's projects.

Some of the suspended projects are disturbingly apposite. One is a study of how ethnic stereotypes fuel conflict between Georgian, Abkhazian, Ossetian and Russian ethnicities, being conducted at Tbilisi State University. An international conference on conflict potential and tourism development, scheduled to take place in early October in the town of Gori, near the border with the disputed region of South Ossetia, is expected to be cancelled.

The crisis has hit Georgia midway into a sweeping reform of its science-funding system. The GNSF, a grant-giving agency set up in 2005 along Western lines, had planned to double its support of national science next year, from US$8 million to US$16 million. But given the costs of post-war reconstruction, the budget increase is now unlikely to materialize, says Jokhadze. Grants for young scientists and travel grants may have to be cut, she says.

The capital Tbilisi remains unscathed. But on 12 August, Russian bombing of Gori set the university ablaze. The full extent of damage there and at universities, hospitals and schools in some of the more remote regions of Georgia is still not clear, because roads and communication channels are blocked or destroyed and journalists have been denied access. In a message posted on the Internet, Zaza Tsotniashvili, rector of Gori University, said that the damage was "extensive" and that the university might have to close.

A ceasefire deal brokered by France officially took effect on 12 August, but military operations continued until 17 August.

International conservation charity WWF estimates that 280 hectares of forest have been burnt in the conflict, and warns that key conservation areas are under threat, including a biodiversity research hotspot in western Georgia. "Since last night, Russian helicopters are dropping bombs in the Borjomi Gorge, in the area of Borjomi-Kharagauli National Park," David Tarkhnishvili, a zoologist at the Ilia Chavchavadze State University in Tbilisi, wrote on 16 August in an e-mail to Italian colleagues. "Multiple fire patches are currently spreading over the area."

Most foreign researchers who have been involved in international research projects have now left the country. Palaeontologist Lorenzo Rook of the University of Florence in Italy, who moved to Georgia in July to participate in an international archaeological study at the Plio-Pleistocene site of Dmanisi in southern Georgia, was evacuated by the Italian embassy on 10 August. The rest of the team, comprising students and scientists from Italy, Spain, the Netherlands and the United States, were also safely evacuated from the excavation site before the Russian airforce began bombing a military base between Dmanisi and Tbilisi.

Rook, who has visited Georgia regularly since 2000, says that the number of young, skilled native researchers has grown remarkably quickly in the country over the past few years, part of a more general scientific resurgence. "I hope Georgian science will recover from the current crisis equally fast, and that we can all go back and resume our work there as soon as possible," he says.

"We very much appreciate Western support," says David Lordkipanidze, director-general of the Georgian National Museum in Tbilisi, who oversees the now-suspended excavations in Dmanisi. "Unfortunately, there is little else we can do other than wait from one day to the next for the Russian occupation of half of our country to end."

Luckily, he adds, the Dmanisi archaeological site, where the oldest human remains outside Africa have been found, is outside the conflict zones. He is optimistic that the open-air museum being constructed there will open before the year's end as planned, with a big celebration and a science festival.

Russia last week began to withdraw its forces from Georgian territory, but military officials say troops will remain stationed in buffer zones around South Ossetia and Abkhazia. On 26 August, the Russian President, Dmitry Medvedev, recognized the two breakaway regions as independent states. ■

Quirin Schiermeier

DOUBLE FIRST FOR LHC
Particles from the new collider's beam detected
www.nature.com/news

CERN

Fresh doubts over *T. rex* chicken link

A claim by researchers to have extracted proteins from a *Tyrannosaurus rex* bone and matched these to proteins found in chickens has been attacked in the same journal that published the original research.

In a withering critique, computational biologist Pavel Pevzner and his colleagues at the University of California, San Diego, say that the protein claim cannot be supported by the analytical data released so far[1].

The original articles, published last year in *Science*, claimed that palaeontologist Mary Schweitzer of North Carolina State University in Raleigh and her colleagues had recovered fragments of collagen from inside a 68-million-year-old *T. rex* femur bone[2] — making the protein 100 times older than the previous collagen record holder, from a mastodon (*Mammut americanum*) that died up to 600,000 years ago.

A linked article described the analyses of the *T. rex* protein samples performed by John Asara, who runs a mass spectrometry research lab at Beth Israel Deaconess Medical Center in Boston, Massachusetts, and his colleagues. Asara was able to match sequences from all the collagen fragments to those of living species including chickens, better defining the evolutionary link between reptiles and birds[3].

But Pevzner calls the article "computationally illiterate". He argues that the mass spectrometry data on the seven proteins recovered are not broad enough to prove a statistically significant match with chicken collagen. Because Asara's team has not revealed all the 48,000 mass spectra data generated, he says, it is impossible to rule out the 'false positives' that are routinely generated by the technique, and so tell whether the protein match is a mere coincidence like "a monkey typing random keys on a typewriter" that by chance spells words.

Spectra of all studied proteins are routinely published as supplementary data to enable scientists to replicate results, but Asara declines to release this data, saying that to do so would open the work to publication by others.

"I'm surprised; I don't understand how they went forward and published unless those data were publicly disclosed," says Richard Smith, head of the mass spectrometry lab at the Pacific Northwest National Laboratory in Richland, Washington.

Asara, whose rebuttal to the Pevzner critique appears in the same issue[4], says his team conducted additional analyses that support their earlier results. "After being forced to go through evaluation of the data, we stand by the article even more so today," he says.

Asara's confidence belies the article's increasing troubles. The researchers withdrew one protein from the work as far back as September 2007, saying that it wasn't statistically significant[5]. Next, in January, *Science* published a technical comment on the article[6], in which 27 authors reported that they could not verify the *T. rex* proteins, to which Asara and Schweitzer again replied[7].

By June, Asara was publicly acknowledging that two other proteins were also not statistically significant. This, despite a short article in *Science*[8] a month earlier, in which Asara, Schweitzer and their colleagues asserted that a comparison of the ancient proteins to existing species — such as crocodile and ostrich — helped to affirm their earlier work.

Even the *T. rex* protein samples have been questioned. On 30 July, Tom Kaye, a research associate at the Burke Museum of Natural History and Culture in Seattle, Washington, asserted that the collagen extracted from the ancient bone was in fact remnants of bacterial slime[9]. Schweitzer told *Nature* that she rejects the evidence, from scanning electron microscope images, because it came from other bones — Kaye says that his team was denied access to the original bone.

With the controversy over their original article unabated, Schweitzer says that she will hold a private meeting in November with invited scientific authorities to develop additional standards for publishing such work. But Pevzner is looking for a different response. "How many technical comments should there be before an article is withdrawn?" he says. ■

Rex Dalton

1. Pevzner, P. A., Kim, S. & Ng, J. *Science* doi:10.1126/science.1155006 (2008).
2. Asara, J. M., Schweitzer, M. H., Freimark, L. M. & Phillips, M. *Science* **316,** 280–285 (2007).
3. Schweitzer, M. H. *et al. Science* **316,** 277–280 (2007).
4. Asara, J. M., Schweitzer, M. H., Cantley, L. C. & Cottrell, J. S. *Science* doi:10.1126/science.1157829 (2008).
5. Asara, J. M. *et al. Science* **317,** 1324–1325 (2007).
6. Buckley, M. *et al. Science* **319,** 33 (2008).
7. Asara, J. M. & Schweitzer, M. H. *Science* **319,** 33 (2008).
8. Organ, C. L. *et al. Science* **320,** 499 (2008).
9. Kaye, T. G., Gaugler, G. & Sawlowicz, Z. *PLoS ONE* **3,** e2808 (2008).

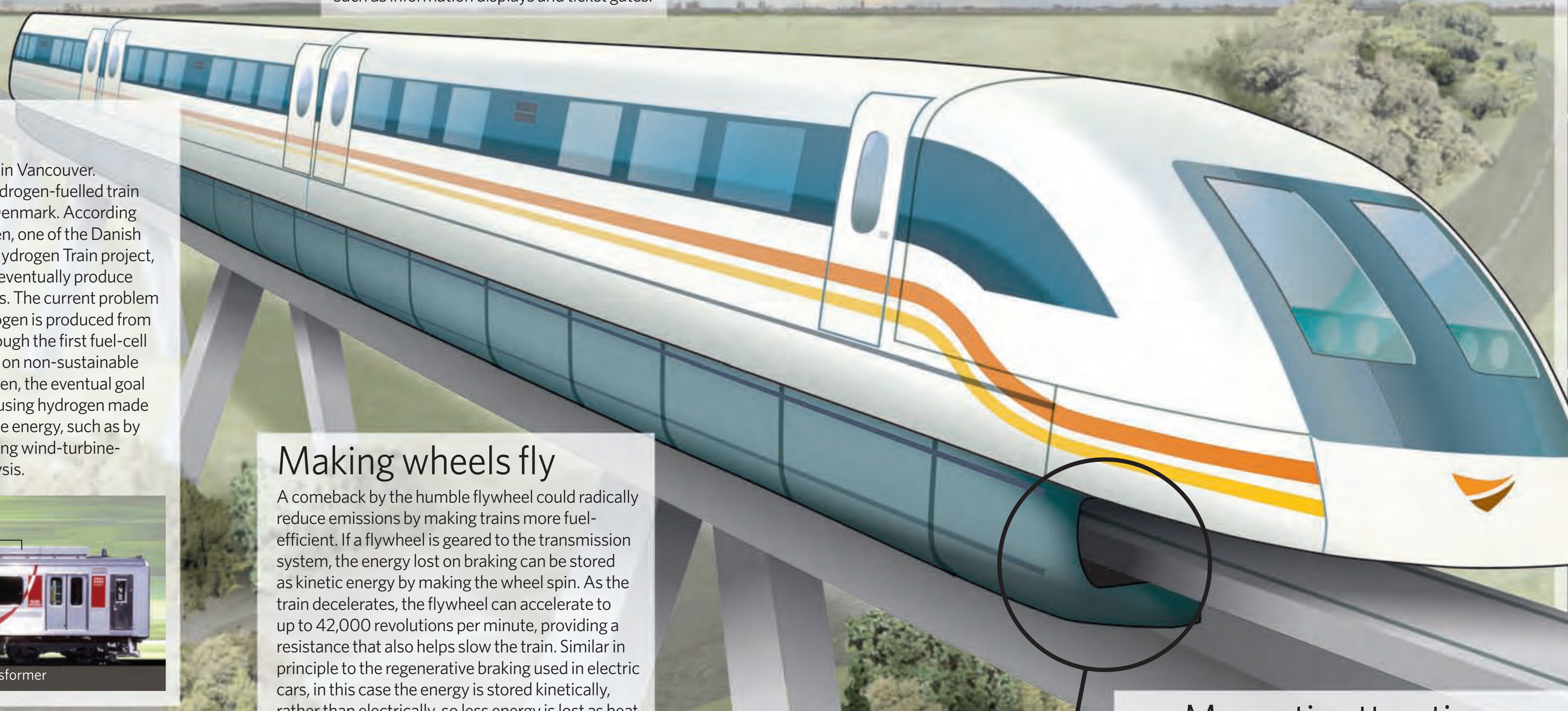
L. PSIHOYOS/CORBIS

E. LAMM/MUSEUM OF THE ROCKIES

Critiques of a *Tyrannosaurus rex* study on protein from its femur bone (left) have been as fierce as the dinosaur.

Do the locomotion

Rail travel produces more than a third less emissions than road transport — even though trains carry 7% of traffic, they emit just 0.2% of the carbon monoxide, 2% of nitrogen oxides and 1% of the volatile organic compounds. Although electric passenger trains are relatively green, most of the world's trains are used for haulage and run on diesel. In the latest of our Future Transport series, **Duncan Graham-Rowe** sees trains switching to a greener track.



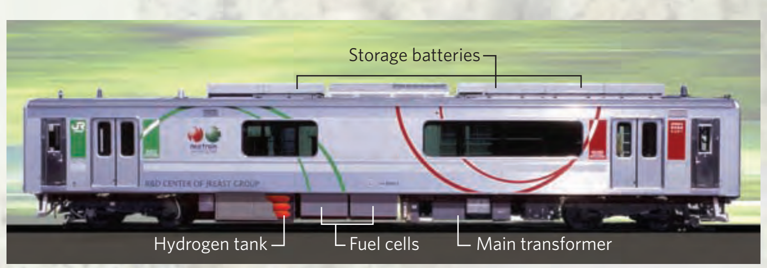
People power

The stations that service the trains also need power, and the East Japan Railway has plans to improve energy efficiency here. It has developed a power-generating floor that uses the vibrations created by commuters walking through the automatic turnstiles to generate electricity. Piezoelectric elements embedded in pads on the floor convert the footfall pressure into electric current. Tests at Tokyo Station showed that the 40,000 or so people who pass over a pad each day can generate 10,000 watt-seconds/day — just enough to light a 100-watt light bulb for 100 seconds. For those who are slightly more demanding of their electricity supply, larger pads will aim to generate enough electricity to power facilities such as information displays and ticket gates.

Running on gas

Several companies are exploring hydrogen fuel cells to power trains. Such 'hydrail' trains are essentially electric but would not need the usual trackside infrastructure. And, unlike diesel trains, fuel-cell-powered trains produce no emissions other than water. Fuel cells produce electric current by combining hydrogen fuel with ambient oxygen using catalytic electrochemical reactions. East Japan Railways has already tested a hybrid fuel-cell train (pictured below) on one of its passenger lines and Canada is planning to develop fuel-cell trains in time for the 2010 Winter Olympics in Vancouver.

Europe's first hydrogen-fuelled train is likely to run in Denmark. According to Claus Torbensen, one of the Danish founders of The Hydrogen Train project, such trains could eventually produce zero net emissions. The current problem is that most hydrogen is produced from methane. So although the first fuel-cell trains have to rely on non-sustainable sources of hydrogen, the eventual goal is to power them using hydrogen made through renewable energy, such as by splitting water using wind-turbine-powered electrolysis.



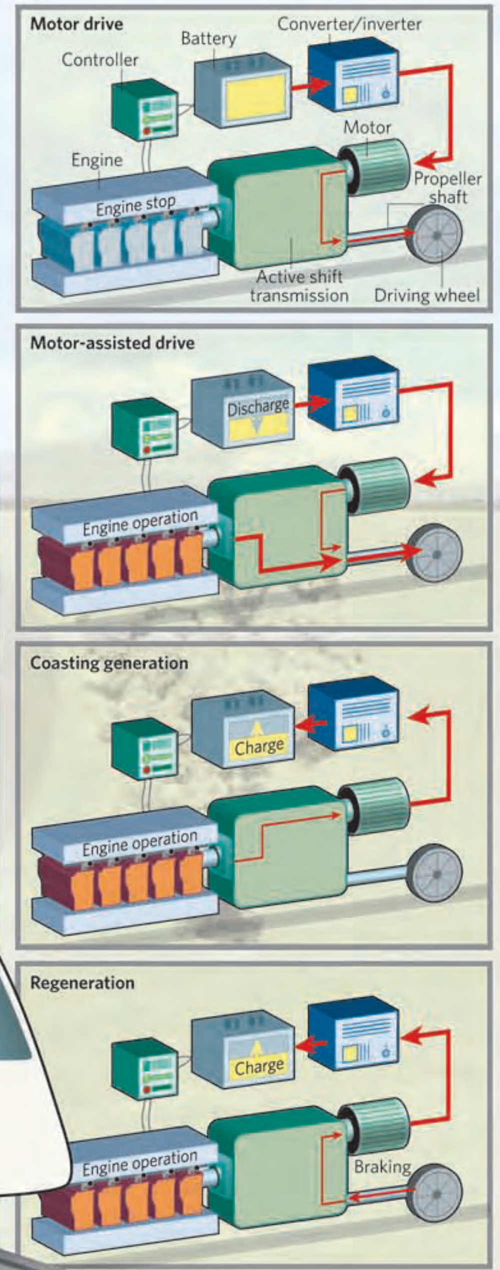
EAST JAPAN RAILWAY CO./AP PHOTOS

Hybrid hopes

As with cars, electrical energy generated by diesel engines and by regenerative braking can be stored and used to drive electric motors, providing additional power when needed, as when accelerating. The Hokkaido Railway Company in Japan operated the first passenger train using this next-generation diesel–electric hybrid technology. It takes the hybrid concept a step further by storing enough electrical energy so that the diesel engine can be turned off during long stops, which saves energy and reduces noise pollution. When setting off again, the train is driven entirely on the electric motor until it reaches about 45 kilometres per hour, when the diesel is brought online again (see graphic).

According to the company, this set-up reduces fuel consumption and emissions by up to 20% compared with conventional diesel–electrics.

RailPower, a Canadian company based in Quebec, has developed a hybrid system that it reckons cuts diesel use by up to 60% and reduces emissions of nitrogen oxides by up to 90%. Its Green Goat trains, which are designed for shunting, use only electric motors to drive the train, with the diesel engines being used to charge the batteries. According to the company, a single Green Goat, carrying out typical railyard operations, will reduce greenhouse-gas emissions by 271 tonnes per year compared with diesel.



SOURCE: HOKKAIDO RAILWAY CO.

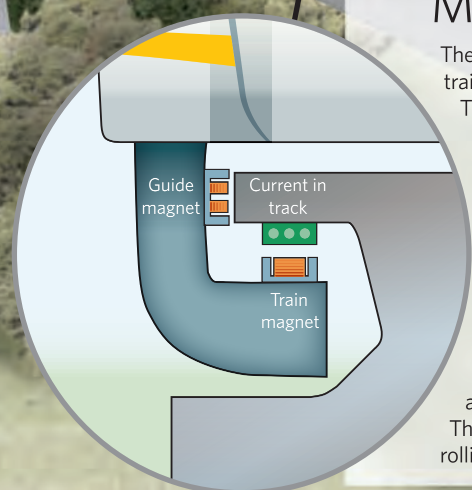
Making wheels fly

A comeback by the humble flywheel could radically reduce emissions by making trains more fuel-efficient. If a flywheel is geared to the transmission system, the energy lost on braking can be stored as kinetic energy by making the wheel spin. As the train decelerates, the flywheel can accelerate to up to 42,000 revolutions per minute, providing a resistance that also helps slow the train. Similar in principle to the regenerative braking used in electric cars, in this case the energy is stored kinetically, rather than electrically, so less energy is lost as heat. This energy can later be released to help drive the train's wheels or to power an electric motor.

Steaming back to the future

Although it may sound unlikely, a return to steam power could reduce some types of emission. The 5AT Project in Britain aims to prove this by building a new-generation steam engine that overcomes the old drawbacks of steam, and improves on the performance and emissions levels of diesel traction engines. Instead of using coal to generate the steam, these new engines will burn diesel or gasoil (a light type of fuel oil).

Although the steam engines will burn more fuel than diesel engines, and so produce more carbon dioxide, because of the way the fuel is burned, almost all of it will be used up, leaving little unburned. So these new steam locomotives will be virtually smoke free, removing emissions of nitrogen oxides almost entirely. "In this respect 5AT should be cleaner than diesel traction," says Chris Newman, one of the members of the 5AT Project team. Initially, such trains are more likely to be used for tourist lines, to replace existing steam engines. But eventually there is potential for 'new steam' to become competitive with diesel haulage trains, says Newman.



Magnetic attraction

The highest-tech solution to driving a train is magnetic levitation, or maglev. This uses a combination of permanent magnets and electromagnets to generate a repulsive force sufficient to lift the train off its supporting rail. At rest, the train is not levitated, but as it begins to move, on wheels, this force increases until there is no longer any contact with the track. The train is propelled by a linear motor that, again using a system of electromagnets built into the track and the train, pushes the train along. The main advantage is that there is no rolling friction, leaving only air resistance and electromagnetic drag to hinder the train. Because of this, maglevs are almost silent at low speeds and can reach cruising speeds of more than 500 kilometres per hour. Maglevs are in operation in China, Japan and Germany, but they are expensive to build because they need an entirely different kind of track. Even so, they produce about a quarter of the emissions of diesel trains, roughly the same as cutting-edge conventional electric trains. And because they are purely electrical, the potential is there to run these trains on entirely renewable energy sources, and so produce zero net emissions.

Death and life beneath the sea floor

Viruses that infect microbes in deep-sea sediments may be a key driver in the world's largest ecosystem and integral to the global carbon cycle, data reveal.

Lytic viruses, which cause infected cells to burst, kill about 80% of the single-celled organisms in the sediment and sub-surface ocean layers, researchers calculate, thereby releasing large amounts of dissolved carbon into the deep seas[1]. Every year this 'viral shunt' releases up to 630 million tonnes of the carbon sequestered by particles sinking into these deep-sea benthic zones, suggesting that viruses should be included in ocean carbon-management models.

The viruses provide a form of population control that may date back to the origin of life on Earth, says bio-oceanographer Roberto Danovaro of the Polytechnic University of Marche in Ancona, Italy, who led the analysis of 232 deep-sea sediment samples with his colleagues. "The viruses kill the microbes and stimulate their growth as well," he says. "It's almost a self-sustaining mechanism."

The viral data are the latest in a series of revelations about the deep-sea ecosystem, which covers 65% of the planet's surface. Around one-tenth of Earth's living biomass exists at the bottom of the ocean, despite cold temperatures, impenetrable darkness and intense pressure. "A few years ago we were assuming the deep-sea ecosystems had no currents, no movements," says Danovaro. "Now we know a lot of material and sediment can be brought up from the deep in a few days."

The nature of the microbes living in and below the sea floor remains hotly debated. It was thought that the majority consists of bacteria. But a study published last week concluded that most cells in the sediment are archaea[2], a similar-looking but distinct form of life.

Previous ocean-floor surveys turned up plenty of bacteria but relatively few archaea. Some studies may have been biased — several were based on DNA extraction and staining methods that may have met with limited success in penetrating the relatively impermeable archaeal cell membrane. Similarly, lipid-profiling methods based on important components of cell membranes called phospholipid-based fatty acids are generally held to be a good marker for living bacteria because these lipids degrade rapidly after the cell dies. Yet many archaea do not make these particular fatty acids.

Kai-Uwe Hinrichs of the University of Bremen in Germany and his colleagues, however, took samples from more than 1 metre below the sediment surface and milled the cells they collected in liquid nitrogen to loosen up their membranes. They also measured a different class of lipid, one that is common in both archaea and bacteria. The researchers discovered an abundance of archaea, finding that at that depth, archaea make up more of the microbial biomass than bacteria do[2,3].

However, results from a few research sites do not reveal what lies on and beneath the entire ocean floor. "We must be careful not to assume that when we find something in a system, it must be true across environments, in different kinds of marine systems," says Mark Gessner, a microbial ecologist at the Swiss Federal Institute of Aquatic Science and Technology in Dübendorf. Indeed, another study published last week surveyed deep-sea hydrothermal vents and found that many of the viruses found there were not lytic, instead being of a type that rarely causes its hosts to burst[4].

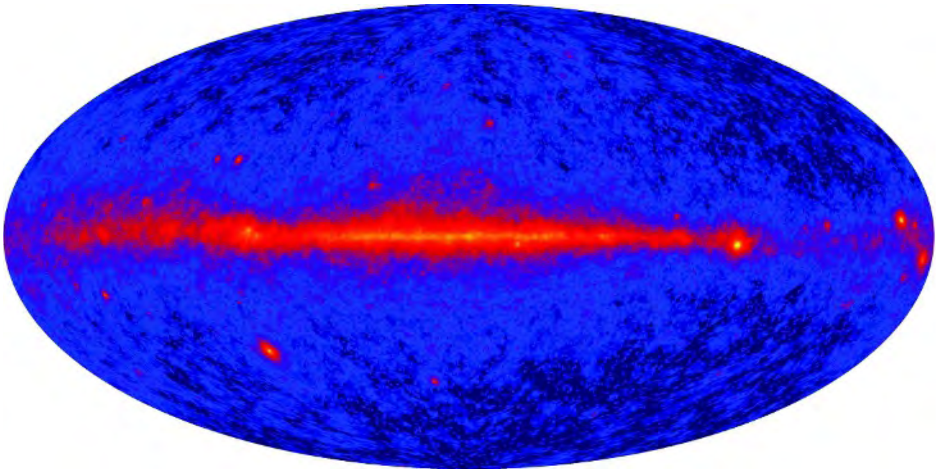
Further technological improvements are needed to get a higher-resolution image of the microbial forest below the sea floor. "We're dealing with very low signal and all of our methods are at the limit of their ability," says Hinrichs. ■

Heidi Ledford

1. Danovaro, R. *et al. Nature* **454,** 1084–1087 (2008).
2. Lipp, J. S., Morono, Y., Inagaki, F., and Hinrichs, K.-U. *Nature* **454,** 991–994 (2008).
3. Biddle, J. F. *et al. Proc. Natl Acad. Sci. USA* **103,** 3846–3851 (2008).
4. Williamson, S. J. *et al. ISME J.* advance online publication doi: 10.1038/ismej.2008.73 (21 August 2008).

SNAPSHOT

New window on the gamma-ray Universe



NASA/DOE/INTERNATIONAL LAT TEAM

NASA's Gamma-ray Large Area Space Telescope (GLAST) was launched on 11 June and turned on its main telescope two weeks later. It was immediately blasted with gamma-rays from blazar 3C454.3 — visible in the lower left quadrant of the telescope's first map of the sky. "It's a good example of the kind of things that are in store for us," says project scientist Steven Ritz of NASA's Goddard Space Flight Center in Greenbelt, Maryland.

Blazars, supermassive black holes that emit tight jets of particles, are just one of the gamma-ray phenomena that GLAST will study as it scans the whole sky every three hours. Other objects clearly visible in the first map are the Geminga and Crab pulsars, above and below the plane of the galaxy at the far right of the image. The map is based on 95 hours of data: it would have taken the Compton Gamma-Ray Observatory, NASA's previous gamma-ray telescope, a year to amass as much data. At the same time as releasing the map, NASA renamed GLAST the Fermi Gamma-ray Space Telescope. ■

Eric Hand

Q&A

Too close for comfort

In early 2002, the US Federal Bureau of Investigation (FBI) asked the American Society for Microbiology to canvas its 43,000 members for information about the 2001 anthrax mail attacks that killed five people. **Nancy Haigwood**, now director of the Oregon National Primate Research Center in Hillsboro, suggested that agents should investigate microbiologist Bruce Ivins, who had been harassing her for more than 20 years. On 29 July, Ivins killed himself as authorities were close to indicting him for the anthrax attacks.

When did you first meet Ivins?
In 1976, he was a postdoc and I was a graduate student in the department of bacteriology and immunology at University of North Carolina (UNC) in Chapel Hill. I may have gone out for a sandwich with Ivins and his wife together.

When did his behaviour first concern you?
In 1982, my house and first-husband's car were vandalized, marked with KKG [Ivins had an obsessive hatred of the Kappa Kappa Gamma sorority, of which Haigwood was a member]. A letter was also ghost-written in my name to a newspaper. I knew it had to be Ivins. I confronted him on the telephone. He denied it. Later, I came to believe that he had stolen a lab book from my UNC locker in 1979. An anonymous note said I could find my missing book in a particular US Post Office Box. What now looks significant is the use of the postal box.

Why did you link Ivins to the anthrax attacks?
I had been getting regular e-mail updates from him. In November 2001, two months after the attacks, he sent out e-mails and photos of himself working with anthrax at the Fort Detrick lab.

Did the FBI ask you to meet with Ivins and wear a hidden recording device?
Yes. I came close to setting a meeting up. I was going to have lunch with him when visiting Washington DC. Ultimately, I couldn't do it. I was afraid. The FBI told me there might be guns involved. I would be surrounded by disguised agents. But the agents thought I would be too nervous to do it.

What are your thoughts about Ivins's death?
I was shocked. There was some relief. You never want someone you know to do

something like the anthrax attacks. But he must have felt cornered — and he was.

Do you have any doubts that Ivins was behind the anthrax mailings?
I have not seen all the evidence. But I have read the October 2007 affidavit to search his house. It tied the mailed anthrax spores to Ivins's flask. I don't think he intended to kill people, but to scare them. Now I know he stood to gain from business from his work.

What is the most significant evidence?
The genetic evidence. And, he intentionally gave the FBI the wrong sample. He covered up an anthrax spill at his lab. He checked out a freeze drier late at night, exactly what you would need to make the mailed spores. All this piles up. But the mailings from a box near the KKG house in Princeton, New Jersey, that was the icing on the cake — a diabolical little twist that is so Ivins.

What should happen next?
I would like to see a scientific paper published on this. I have no reason to doubt the FBI; I met a half-dozen agents, I got to know two very well. But the scientific community needs to see all the data. ■

Interview by Rex Dalton

Rector sacked in Austrian stem-cell scandal

A fierce academic dispute in Austria reached fresh heights on 21 August when the Medical University of Innsbruck's council dismissed the rector, German immunologist Clemens Sorg, without notice.

In a letter to friends and colleagues in Austria and Germany that was later leaked to the press, Sorg had criticized Austria's academic establishment for its laxness in a case of scientific misconduct at the university's urology department.

Earlier in August, a report by the Austrian Agency for Health and Food Safety found that a clinical trial involving stem cells, run by urologist Hannes Strasser at the university, had serious flaws (see *Nature* 454, 922–923; 2008). Last week the agency announced that it will hand the report over to public prosecutors.

But the seven-strong university council says that Sorg's "public attacks" against his host country were "unjustified". It accuses Sorg of a "serious breach of duties", including violating official secrecy. In a copy of his letter that appeared in the Austrian daily *Tiroler Tageszeitung*, Sorg writes that Austria's highest "networks" were attempting hush up what he calls a "medical scandal of unprecedented scale". Sorg says that he intends to sue the Medical University of Innsbruck.

Nuclear group to rule on Indian trade

The Nuclear Suppliers Group (NSG) will soon decide whether India can trade nuclear fuel and equipment with the group's members. The collaboration met in Vienna last week, and is due to meet again in September.

In the 34 years since its first test, India's nuclear programme has been limited by the nation's refusal to put its name to the nuclear non-proliferation treaty, which all members of the NSG have signed. India wants access to international nuclear trade without restrictions on the growth of its nuclear weapons programme.

A nuclear deal between India and the United States is poised for approval by the US Congress on the condition that the NSG agrees to India's request. Any of the NSG's 45 member countries can veto the application.

NIH promises funds for cheaper DNA sequencing

The US National Human Genome Research Institute (NHGRI) is ploughing more than $20 million into new genetic sequencing technologies.

A series of grants announced on 20 August are the latest step in the institute's drive to bring down the cost of sequencing. The money will support projects such as the development of nanopores. These structures could, it is hoped, identify DNA bases threaded through them from variations in the bases' ionic or electrical properties. The largest grants will go to Daniel Branton and Jene Golovchenko of Harvard University, who are developing this technology, and Mostafa Ronaghi of Illumina in San Diego.

Jeffrey Schloss, director of the NHGRI's technology-development programme, says that the institute's goal of a $1,000 genome by 2014 "is still realistic".

Inquiry launched into Indian drug trials

The All India Institute of Medical Sciences is investigating child deaths in several drug trials.

M. VATSYAYANA/AFP/GETTY IMAGES

The All India Institute of Medical Sciences (AIIMS) last week launched an inquiry into the deaths of 49 children under 12 years of age who have died in drug trials at the New Delhi institution since January 2006. But questions are being raised over whether the inquiry is necessary on scientific grounds.

The death rate in the 42 separate clinical trials was just over 1%. The trials involved a total of 4,142 young patients, 2,728 of whom were under a year old. Among the drugs being tested were the cancer drug rituximab, as well as olmesartan and valsartan, which lower blood pressure.

"The deaths were due to sickness only. All drugs used were of proven safety," Shakti Kumar Gupta, the head of hospital administration at AIIMS, told *Nature*. The overall mortality rate at the institute, which treats seriously ill patients from all over India, is 3.6%.

Cracks spotted in Greenland's glaciers

A team at Ohio State University in Columbus has reported a large crack in the longest floating glacier in the Northern Hemisphere. If the crack becomes a break, 160 square kilometres will be lost from the Petermann glacier in northwestern Greenland. This would greatly exceed the glacier's last major ice loss, which saw 86 square kilometres break away in 2000–01.

Glaciologist Jason Box and his colleagues spotted the crack on 3 August. In addition, their daily checks of satellite images of Greenland's glaciers have revealed 29 square kilometres of ice loss from Petermann between 10 and 24 July.

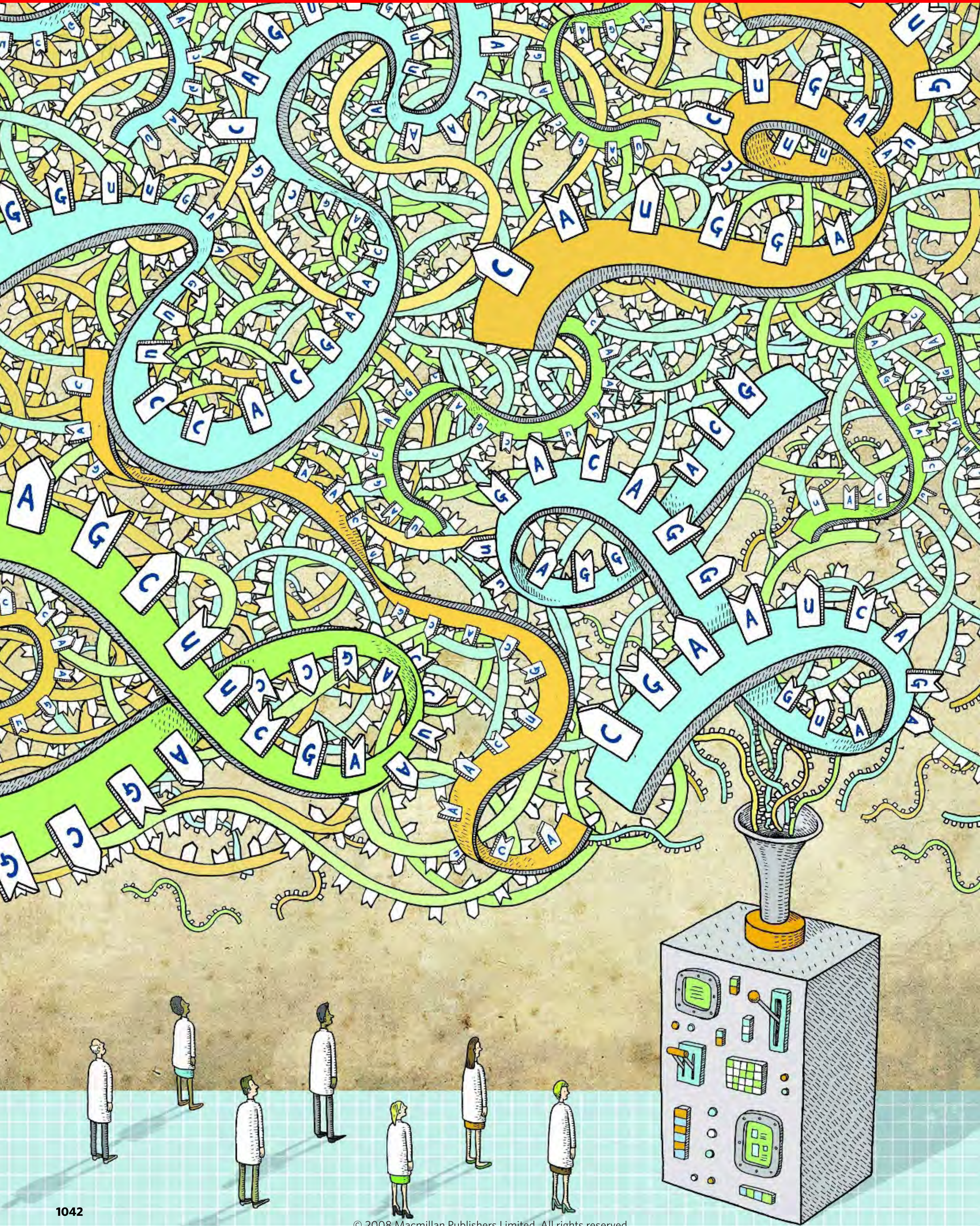
They have also seen 10 square kilometres of ice loss since the end of the last melt season from the Jakobshavn glacier further to the south, which is the world's fastest-retreating glacier.

NASA

Orion crash-landing leaves NASA hunting for clues

Photos released last week show the remains of NASA's next-generation spacecraft, which made a hard landing after a critical parachute system failed to deploy properly.

A mock-up of the Orion crew vehicle slammed into the US Army's Yuma Proving Grounds in Arizona on 31 July after it "dropped faster than intended", the agency said (pictured). The Orion capsule is set to replace the space shuttle, which is due to retire in 2010. A report released 21 August by the US National Academy of Sciences warned that only 5 of NASA's 22 lunar research projects are on track and risk-free.



THE PRODUCTION LINE

If more than 90% of the genome is 'junk' then why do cells make so much RNA from it? **Anna Petherick** goes in search of some answers.

HOTAIR is a molecule with a future. Created from a DNA sequence on human chromosome 12, it affects genes on chromosome 2, apparently working as part of the system that enables skin cells to tell where on the body's surface they are, and thus what they should be doing.

Beyond these specifics, HOTAIR may also serve as a model for understanding a whole slew of similar molecules, the existence of which was not even dreamed of ten years ago and the function of which — if any — is still hotly debated. HOTAIR stands out because it is a long piece of RNA that doesn't encode a protein but still does something biologically important[1]. "HOTAIR was a gem in a sea [of long RNAs]," says John Rinn, a genome biologist who discovered the RNA while working at Stanford University in California. "It told us little about what the bulk of these things are doing. For that, we can't even see a common trend."

> "Many transcipts are made that we don't understand. We still don't know what those transcripts do, if anything." — Ewan Birney

It is hard to comprehend the upheaval that RNA has been causing in molecular biology over the past few years. Once viewed as a passive intermediary, it was thought to faithfully carry genetic messages from the DNA sequence to the protein-making machinery, where things were made that actually got things done. Biologists were comfortable in the knowledge that only 1–2% of the human genome made protein-coding RNA in this way, and most of the rest was filler. So when, in 2005, geneticist Thomas Gingeras announced that some cells churn out RNA molecules from about 80% of their DNA, he astonished scientists attending the Biology of Genomes meeting at Cold Spring Harbor Laboratory in New York. Why should cells bother with so much manufacturing if, as it seemed, such a tiny fraction was involved in the important business of protein-making?

Over the past three years or so the case for this 'pervasive transcription' has strengthened. The phenomenon has now been ascribed to mice, fruitflies, nematode worms and yeast. These studies, and Gingeras's original reports, came from microarrays — a technology that relies on the tendency of nucleic acids to find their complementary cousins in a solution. Gingeras works for the microarray manufacturer Affymetrix in Santa Clara, California. But not everyone has been persuaded of the extent of pervasive transcription, in part because microarrays are subject to background 'noise'. Even using no RNA, control chips will give off some signals, and results can be a matter of interpretation.

For anyone who still doubts that the genomes of nucleated organisms are first and foremost RNA machines rather than protein-coding ones, sequence data are starting to provide "ultimate information", Gingeras says. There is something about the nitty gritty of nucleotide sequences that is enticingly reassuring to molecular biologists. New sequencing machines that can stream out data many times faster than their predecessors have made the mass sequencing of cellular transcripts possible.

In 2008, this process was completed for two species of yeast[2,3] using machines made by Illumina, based in San Diego, California. The results broadly agree with the microarray findings, showing transcription from 74% of the genome of brewer's yeast (*Saccharomyces cerevisiae*) and 90% from that of fission yeast (*Schizosaccharomyces pombe*). Gingeras and other researchers are now working to sequence all the RNA produced by 44 kinds of human cell as part of the Encyclopedia of DNA Elements (ENCODE) project, which aims to identify all the functional parts of the human genome. At that point, any remaining sceptics will be able to overlay the many thousands of different human RNAs onto DNA regions from whence they came. At the end of this process, the covered regions will be those that give rise to RNA — and the uncovered ones, probably just a few naked holes.

All this transcriptional accounting has hastened an already heady RNA rush. Even before the pervasive nature of transcription became clearer, molecular biologists had begun to trot out new classes of RNA molecules that are responsible for important happenings in cells. Thrust farthest into the limelight are the microRNAs (miRNAs), which stop the production of certain proteins, but they have been joined by a growing number of other RNA families, such as small nucleolar RNAs (snoRNAs) and Piwi-interacting RNAs (piRNAs), with vital roles in cellular and developmental processes — vital enough to earn the DNA that encodes them the label 'RNA genes'.

The long and the short of it

On the whole, the established classes of RNAs are short molecules, around 20 or 30 nucleotides in length. The non-coding RNAs that Rinn has been championing run to 200 or even 10,000 bases apiece. The issue at the moment is whether, among this bounty of long RNAs, researchers will find anything as biologically meaningful as the shorter RNAs have proved to be. HOTAIR shows that some such molecules have function — but is it the exception or the rule? "It's controversial whether these are mostly just noise or regulatory function," says Jürg Bähler of the Wellcome Trust Sanger Institute in Cambridge, UK, who led one of the yeast RNA sequencing projects.

Those who doubt the importance of RNA bemoan their logical problem: it is impossible to prove lack of function. Even when an important cellular job does get pinned on a long RNA, as it did for HOTAIR, the doubters worry that it is too tempting to extrapolate across the board.

K. MAR

John Rinn may have found a new class of long RNA genes.

P. HACKETT

Ewan Birney, a bioinformatician at the European Bioinformatics Institute and one of the leading scientists in ENCODE, says that the debate now is about what proportion of long RNAs serve a purpose. "I used to be a much stronger sceptic three to four years ago," he says. "Now I'm accepting that transcription is pretty complicated and that many transcripts are made that we don't understand. Where I still have some scepticism — what we still don't know — is what those transcripts do, if anything."

John Mattick, the director of the Centre for Molecular Biology and Biotechnology at the University of Queensland in Brisbane, Australia, has no such qualms. He is a long-time advocate of non-coding RNA's importance. The doubters, he says, "keep regressing to the most orthodox explanation [that the long RNAs are junk]. But they can't just sit on their intellectual backsides and tell us to prove it." But prove it is just what researchers are starting to do, with a growing number of examples that showcase these molecules' capabilities.

The idea of long non-coding RNAs is not new. Xist, the most famous example, was discovered in 1991. Its 17,000 nucleotides can be found in almost every cell of mice and humans, where it obviates gene expression along an entire X chromosome. Because females have two Xs to their male (XY) counterparts' one, they use Xist to switch off the extra X and compensate for the disparity.

Varied roles

Xist RNA is transcribed from the chromosome it mutes, and coats it along its length. No one really knows exactly how it attaches and what makes it so effective at gene silencing. What is clear, however, is that part of the molecule attracts chromatin remodelling complexes — enzymes that turn genes on and off by tinkering with DNA's packaging. Get enough of these complexes together, and it seems that you can turn off a whole chromosome.

Over the past few years, the RNA field has compiled a brief list of other long non-coding RNAs. Many of those that have been studied control the activity of protein-coding genes. As the pace of these discoveries has picked up, they have revealed that long RNAs can control genes in a surprising variety of ways, from both near and far, and that their function is not necessarily dependent on the exact sequence of the RNA, as it is when RNA is coding for proteins. This suggests that scientists have only begun to appreciate what RNA is capable of.

> "Either there's a hell of a lot of developmentally regulated transcriptional noise, or these RNAs are sending signals into the system."
> — John Mattick

In one example published last year, molecular biologist Igor Martianov and his colleagues at the University of Oxford, UK, studied the human gene for dihydrofolate reductase, an enzyme involved in biochemical syntheses that has two 'on' switches for protein production. They discovered that the first of these switches actually triggers the manufacture of a 583-nucleotide-long RNA molecule, and that this RNA directly interferes with the second switch. When this happens, the enzyme is no longer made[4].

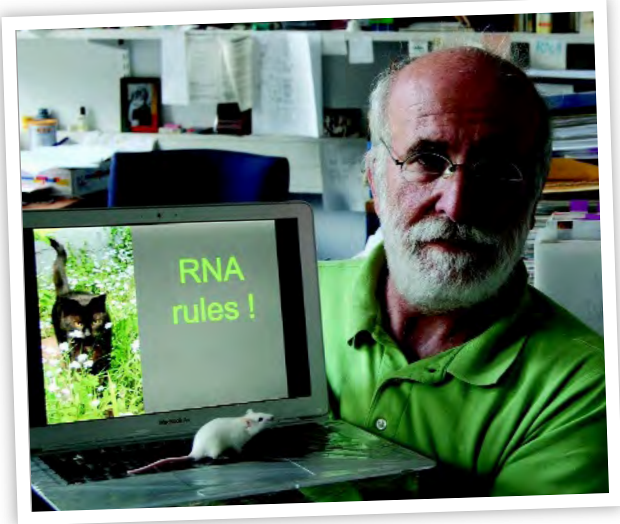
Working in a very different way, a long RNA called NRON seems to travel to the cytoplasm in order to influence the expression of protein-coding genes. Several thousand nucleotides long, NRON polices the trafficking of a transcription factor from the cytoplasm into the nucleus of the cells where it is active[5]. By doing so, it seems to control the transcription factor's activities, which include regulating T cells' immune response.

When Rinn discovered HOTAIR, it reinforced the idea that RNAs could be shuttling around the genome doing important jobs. Rinn was studying skin-cell lines cultured from the finger, foot, foreskin and eight other sites on the human body, trying to find out how these cells know their position.

HOTAIR, which stretches for nearly 2,200 nucleotides, is produced from within a cluster of the *HOX* genes that specify an early embryo's head end and foot end, as well as the order of the body segments in between. When Rinn found that this RNA affects the output of genes on chromosome 2, it was the first time such a cross-chromosome influence had been found. When he lowered levels of the RNA molecule, the activity of *HOX* genes on chromosome 2 jumped, and foreskin cells started behaving in an unusual way[1].

Rinn initially wanted to name the molecule STAR1. The acronym for 'Suz-Twelve Associated RNA' refers to the enzyme that ferries this molecule from one chromosome to another, and the number one reflected Rinn's optimism that there are likely to be more STARs. But Rinn's lab partner, Howard Chang, wanted a "more humbling" name, Rinn says, and they settled on HOTAIR instead (for *HOX* antisense intergenic RNA). "Howard was right, but I think we are still both in search of more stars, not hot air," says Rinn, now at the Broad Institute in Cambridge, Massachusetts.

As Rinn has said, there is a vast sea of long RNAs out there. The ones with functions already ascribed to them comprise just a minuscule fraction, and those seem to be regulating genes by very diverse means. To many, this lack of common function infers that science has only scratched the surface of the diversity of long RNAs. The massive scale on which transcription is taking place could be the least of biologists' problems compared with its mind-boggling functional complexity. What is needed, researchers say, is more data to show that RNAs do something useful on the genomic scale — but those data are proving remarkably difficult to collect.



L. E. G. BROSIUS

Jürgen Brosius: an advocate of non-coding RNA.

One problem, when it comes to surveying RNA's usefulness, is that sequence does not provide any simple indicator of function. The sequence of non-coding RNA is not conserved between species in the same way that it is for protein-coding genes. If a sequence is doing something important for an organism because of the protein it codes for, then evolution is likely to have kept that region more constant across related species compared with any average stretch. But the same isn't true of RNA, which does not necessarily pair up with a complementary nucleotide sequence at all. Xist is not conserved in this way, nor are any of the other non-coding RNA stars along their full lengths.

Another way to seek evidence of function en masse is to get rid of long non-coding RNAs and watch how animals cope. But such an experiment may produce only subtle changes in an organism as a whole, and could still miss the importance of a transcript. "I think the cell will use these transcripts at very different times and in very different cell types and conditions," Gingeras says. "You may need to see them in a very specific context to see the function."

That is what Jürgen Brosius of the University of Münster, Germany, and his colleagues found when they removed a 150-nucleotide

RNA from mouse neurons, where it is normally transported down the cellular fingers that communicate with other cells[6]. The engineered animals looked and acted more or less the same as the control animals — but Brosius says that on close inspection they weren't as inquisitive and had unusual exploratory behaviours. Such activity might be lethal in the wild, Mattick says, "but it was affecting their behaviour in ways that were far too subtle to be assessed in a cage".

In search of function

If slicing out non-coding RNA doesn't often reveal its function, then perhaps looking at its lifespan will. This vein of thinking brings a potentially bigger blow for RNA's believers than the knockout studies: the possibility that cells are destroying long RNAs almost as fast as they are making them. Studies in yeast have shown that many long RNAs seem to be so rapidly gobbled by the nuclear exosome — a protein complex that degrades RNA — that it is hard to imagine them having any function at all. Some are labelled for destruction as soon as they peel away from their DNA blueprint. David Tollervey, who studies RNA processing at the Wellcome Trust Centre for Cell Biology in Edinburgh, UK, says that long RNAs could have almost-instant effects or cells might be making many long RNAs merely to show that they've done so. In other words, the point of the exercise might be transcription itself, rather than the transcript.

There are already known examples in which RNA production seems more important than the actual product. In 2004, Fred Winston and his colleagues at Harvard Medical School in Boston, Massachusetts, studied a 551 nucleotide RNA called SRG1 that is made by brewer's yeast[7]. It switches on and off the adjacent gene *SER3*, which helps make serine (an amino acid that the yeast needs to be healthy). But in this case it is the process of making the non-coding RNA that regulates *SER3*, rather than the RNA itself. The trick here is that the DNA sequence from which SRG1 is transcribed runs through the on switch for *SER3*. So when a yeast cell is manufacturing a lot of RNA for SRG1, it blocks access to the *SER3* switch. This is what happens when the yeast sits happily in a flask of rich medium and has no need to generate its own serine.

In his transcriptional surveys of humans, Gingeras has shown that about three times as many transcripts carry a molecular label for rapid destruction than do not carry one. But Gingeras thinks that these apparently doomed RNAs still do more for cells than just getting made. When a map of pervasive transcription is overlaid with a map of short non-coding RNAs, such as microRNAs, the two overlap[8]. Gingeras thinks that the short RNAs are frequently embedded within the longer transcripts, and then excised.

Over the past few years, Mattick has been gathering other circumstantial evidence that long RNAs have widespread function. In a paper published in January, he and his team examined 1,328 non-coding RNAs whose expression patterns had been mapped in the Allen Brain Atlas, but the functions of which were unknown. The team found that nearly two-thirds of these molecules were produced in specific regions of the mouse brain — in certain cell types or in specific parts of neurons[9]. More recently, Mattick's team identified 174 non-coding RNAs that are expressed in mouse embryonic stem cells in a decidedly selective manner, either correlating with the cells' capacity to develop into any other cell type or with particular events along the path to specialization[10]. "You've only got two alternatives," Mattick concludes. "Either there's a hell of a lot of developmentally regulated transcriptional noise, or these RNAs are sending signals into the system."

"Transcripts will be used at very different times and in very different cell types and conditions." — Thomas Gingeras

This approach should gain more steam as part of ENCODE. The next-generation sequencers have been chugging away since the end of last year, and in 2009 should lay out the sequences of all the RNA molecules manufactured by two types of human cell. When the project eventually delivers transcriptomes for all 44 cell types, it will allow a closer analysis of when different sorts of human cell make different long RNAs and help infer something about their function.

T. GINGERAS

Thomas Gingeras described 'pervasive transcription'.

As for Rinn, he already has evidence that non-coding RNAs are so much more than hot air. In May, at this year's Biology of Genomes meeting, he presented work suggesting that there are as many as 2,000 long non-coding RNAs in human cells that shoulder biological responsibilities on a par with those of HOTAIR and that may therefore earn the status of RNA genes.

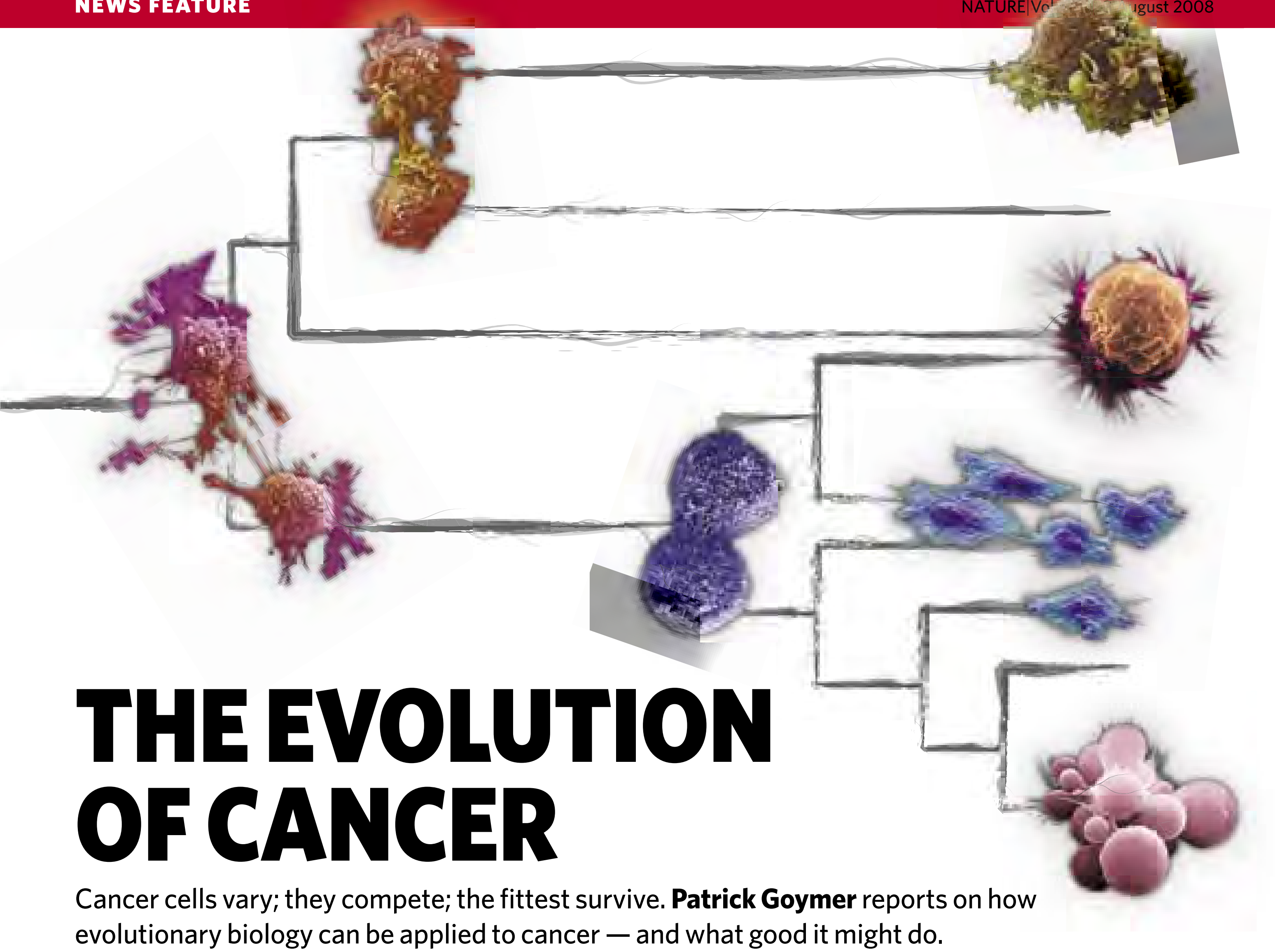
To find these, Rinn and Manolis Kellis, a computational biologist also at the Broad Institute, searched for sequences that are conserved as might befit a working stretch of RNA. They assumed that much of an RNA molecule's function depends on the three-dimensional architecture that the single-stranded molecule folds into. This, rather than the precise sequence of nucleotides, is what evolution will have worked to preserve. This means that an A can become a T, for example, as long as the T to which it anneals when the molecule folds switches in turn to an A and providing that the overall shape of the molecule is unchanged.

Using these types of bioinformatic rules, the team pulled out probable RNA genes. For a sample of these, they took a stab at predicting function and then tested whether the RNA's production was induced by certain cellular pathways. Many of them were. If their results hold up, Rinn and Kellis will have discovered the first large class of long RNA genes. "These RNAs could have functions as diverse as those of protein-coding genes," Rinn says. And it is not such a stretch to think that they could rival the 20,000-odd protein-coding genes in number, if there are other, as yet unidentified groups of long RNA genes out there.

That still leaves a lot of the transcriptional hairball unaccounted for, and it is possible that much of it is still noise. "With all this pervasive transcription," Rinn says, "the problem to working out whether most of it is functional or not has been that people simply haven't known where to start." Now, perhaps, they do. ■

Anna Petherick is *Nature*'s Research Highlights editor.

1. Rinn, J. L. *et al. Cell* **129,** 1311–1323 (2007).
2. Nagalakshmi, U. *et al. Science* **320,** 1344–1349 (2008).
3. Wilhelm, B. T. *et al. Nature* **453,** 1239–1245 (2008).
4. Martianov, I., Ramadass, A., Barros, A. S., Chow, N. & Akoulitchev, A. *Nature* **445,** 666–670 (2008).
5. Willingham, A. T. *et al. Science* **309,** 1570–1573 (2005).
6. Skryabin, B. V. *et al. Mol. Cell. Biol.* **23,** 6435–6441 (2003).
7. Martens, J. A., Laprade, L. & Winston, F. *Nature* **429,** 571–574 (2004).
8. Kapranov, P. *et al. Science* **316,** 1484–1488 (2007).
9. Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. *Proc. Natl Acad. Sci. USA* **105,** 716–721 (2008).
10. Dinger, M. E. *et al. Genome Res.* doi:10.1101/gr.078378.108 (2008).



THE EVOLUTION OF CANCER

Cancer cells vary; they compete; the fittest survive. **Patrick Goymer** reports on how evolutionary biology can be applied to cancer — and what good it might do.

The oncology clinic isn't a field site where one might expect to find an evolutionary biologist. But within the complex ecosystem that is the human body, tumours grow, mutate and face diverse selective pressures as they change and react to their environment. Over hundreds of generations, cells can acquire mutations that promote their errant growth and survival. This makes for diversity both between cancer types and within an individual tumour. But just as species have evolved convergent similarities, cancers too have common themes and steps along their developmental paths. If properly directed with evolutionary theory in mind, treatments might become more effective (see 'Targeting what isn't there').

Tony Green of the University of Cambridge, UK, and his colleagues have looked at evolutionary processes in myeloproliferative disorders — overgrowths of blood-producing bone-marrow cells that can become cancerous. Changes to the *JAK2* gene play an initiating role in these disorders, allowing the cells to bypass their growth-control mechanisms. Green and his colleagues began to study these mutations as the disorders progressed, in some cases, towards a cancer of the white blood cells called acute myeloid leukaemia, or AML. As expected, the *JAK2* mutation arises often and early in myeloproliferative disorders because of the growth advantage it confers on cells. But three of four individuals who went on to develop AML no longer had the mutation[1]. "This was a surprise," says Green. "An initiating mutation was not present in the more evolved state."

Did cancer cells that had acquired *JAK2* mutations lose them over time as other mutations and physiological changes took over the controls of the disease? Or were the *JAK2* mutants outcompeted by other cells taking advantage of the changing environment within the cancer-afflicted individuals?

Green stumbled across this evolutionary parallel, but some scientists specialize in comparing the similarities between changes to a cell in the body and the evolution of organisms within an ecosystem. As more information about cancer genetics accrues, the importance and usefulness of this evolutionary analogy is becoming clear.

Science has been looking for commonalities in cancer, and several large-scale projects aimed at sequencing the genetic changes in different cancers have in their earliest stages revealed what many feared. The main feature of cancer, says Bert Vogelstein of Johns Hopkins University in Baltimore, Maryland, is its complexity and heterogeneity. Most mutations found in cancer are rare. "There are a few genes that are commonly mutated — we call these the mountains — but the landscape is dominated by hills," says Vogelstein. Evolutionary theory, in conjunction with the sequencing of cancer genomes, could help map that countryside more quickly.

Diversity breeds success

Peter Nowell of the University of Pennsylvania in Philadelphia first developed the idea of cancer as a Darwinian process in 1976 (ref. 2). Cancer is known to occur because of the stepwise accumulation of mutations in certain cells of the body. Nowell added to this the population-genetics idea of clonal expansion, in which cells that have a mutation to make them grow faster or survive better produce more offspring than surrounding cells without the mutation.

Carlo Maley of the Wistar Institute in Philadelphia sees the diversity of cancer cells as key to understanding their resistance to drug treatment. "One thing that is surprising is that the multidrug therapies in cancer haven't worked nearly as well as they have in HIV," he says. "That seems to me to be a basic evolutionary question that should be addressed and is at the heart of why we haven't been able to cure cancer."

Maley has been applying evolutionary theory to a condition called Barrett's oesophagus, which can progress to become cancer. As surgical treatment for Barrett's oesophagus is extremely risky, standard medical practice is to monitor cells in the oesophagus for signs that they have started to progress towards cancer. Maley uses biopsy samples to track the evolution of the disorder, testing each biopsy for changes in specific genes such as *CDKN2A* and *p53*. His group has found that in the early stage of the disorder, individuals with diverse populations of cells harbouring different mutations are more likely to develop cancer[3]. This might be because the body is struggling to defend itself against more kinds of attacks. Maley uses methods borrowed from ecology to measure the diversity and make predictions about progression.

Quick and easy

Perhaps the most important advance in cancer biology has been cheap and fast DNA sequencing. The technology that allows researchers to sequence the genomes of hundreds of species, and of individual humans, is now being applied to the genomes of tumours. Knowing the genome sequence of a cancer cell allows scientists to look in detail at how a tumour has evolved from the normal cells of the body — which genes have mutated, how much of the original genome has been lost or duplicated, and whether the evolutionary process has unfolded similarly in each individual case.

Several large-scale projects are taking this approach, including the Cancer Genome Project, which is sequencing protein-coding genes in cancer cells to look for mutations; the Cancer Genome Anatomy Project, which is looking at levels of gene expression in cancer cells; and the Cancer Genome Atlas, which is looking at various types of genomic alteration in specific cancer samples.

But cancer genome sequences aren't by themselves going to explain the evolutionary process of tumour development. In fact, Maley and Green point out that the sequences provide only 'snapshots' of the evolutionary process, so further work is needed to fill in the gaps, such as the order in which the mutations appear. And current technology means that

PHOTOLIBRARY.COM; L. M. DE LA MAZA/PHOTOTAKE/NEWSCOM; A. CAVANAGH/WELLCOME IMAGES

Targeting what isn't there

Drug developers have long had cancer-causing mutations in their sights. But cancer cells invariably evolve ways to become resistant to drugs and ensure survival. Alexander Varshavsky at the California Institute of Technology in Pasadena suggests that drugs should be targeted at something that arises in the cell's evolution that is not so easily side-stepped — deletions of DNA segments[6].

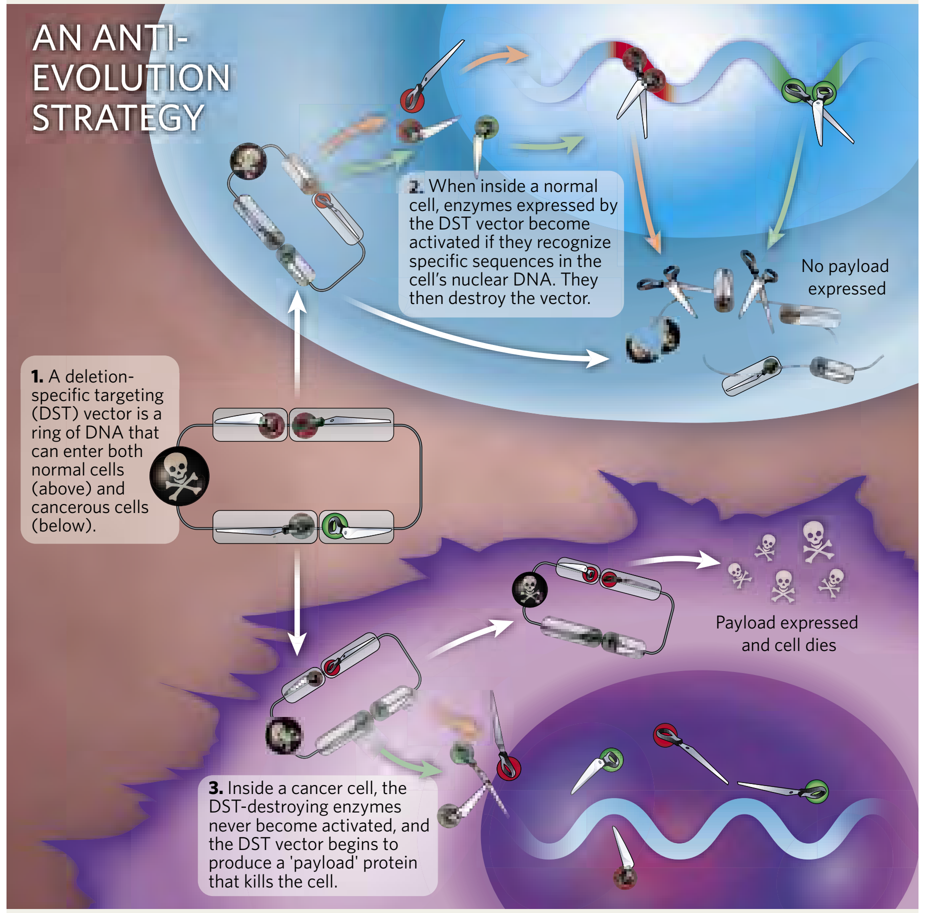
A fundamental principle of evolutionary genetics is that once a gene is lost it is very unlikely to be regained — a phenomenon known as Muller's ratchet. Varshavsky thinks that chance deletions occurring early in a tumour's development could be a hallmark of that tumour whatever course its subsequent evolution takes.

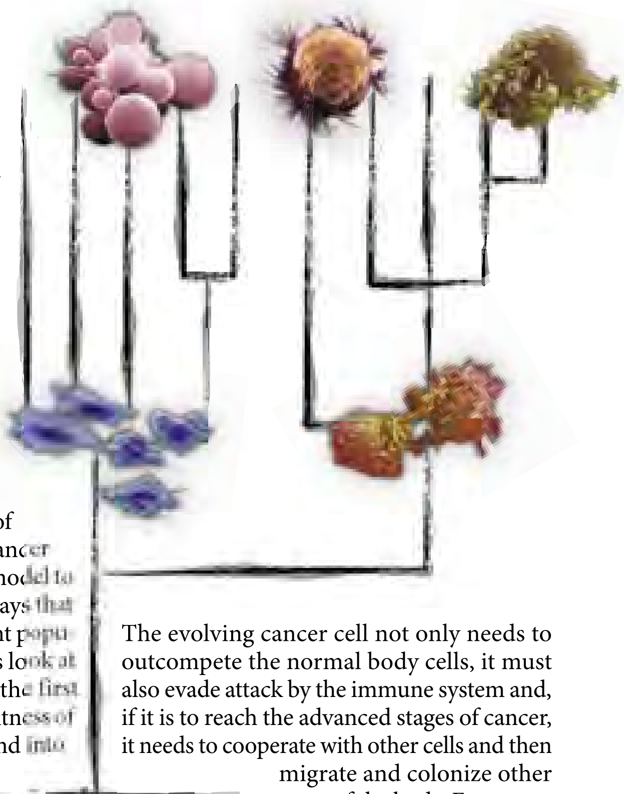
Varshavsky envisages a deletion-specific targeting (DST) vector — a ring of DNA that encodes a cell-killing 'payload' protein and fail-safe enzymes that will destroy the vector when they recognize specific sequences of DNA within the cell. In normal cells, the fail-safe enzymes become activated and destroy the vector before it has a chance to release its payload. Because the specific DNA sequences are missing in cancer cells, the enzymes never become activated and the vector begins to express its deadly payload (see graphic).

Caveats abound. The diverse and shape-shifting nature of cancer means that identifying effective deletion sequences will be difficult. Carlo Maley of the Wistar Institute in Philadelphia, Pennsylvania, cautions that cancer has a knack for overcoming obstacles, including deadly payload proteins. Moreover, the strategy is predicated on gene-delivery techniques that have not yet been proved in cancer.

Still, experts are excited. "It's a brilliant idea," says Bert Vogelstein of Johns Hopkins University in Baltimore, Maryland, "because it exploits the Achilles heel of cancers. Deletions are likely to be present in every cancer."

Varshavsky hopes that the US$1-million Gotham Prize, which was awarded to him last year, will allow him to develop his blueprint into a clinical reality. "I'm committed to implementing the deletion-specific therapy strategy and/or its descendants, taking them as far as they can go. All the way to patients, I hope." **P.G.**





the genome sequences are actually an 'average' sequence taken from a heterogeneous collection of tumour cells, whereas much of the interesting detail is in the differences between individual cells within a tumour — after all, variation is the basic stuff on which natural selection acts.

The need for a sophisticated evolutionary understanding of cancer led Vogelstein to team up with biologist and mathematician Martin Nowak from Harvard University in Cambridge, Massachusetts. Nowak has applied his modelling ideas to problems as diverse as the evolution of HIV, altruism and the politics of climate change. Cancer, he says, "is just like any other evolutionary process, but it's even simpler. Because of this we can ask much more complicated questions."

Sequencing the genomes of cancer cells, says Nowak, can "help us get quantitative data to calibrate our evolutionary models". From Vogelstein's data on sequence variation between individual colorectal cancers, Nowak could predict when malignant tumours would arise from benign ones and when they would metastasize, or spread to other parts of the body[4]. He found that malignant tumours do not mutate more frequently than normal cells, as is often thought. Instead, it is the evolutionary context in which these mutations occur that matters.

Nowak's former student, Franziska Michor, now at the Memorial Sloan-Kettering Cancer Center in New York, is interested in developing her mentor's approach of modelling the process of cancer evolution. The roots of this approach go back half a century. In the 1950s, Richard Doll at the University of Oxford, UK, found that solving equations containing terms for growth and mutation rates allowed him to predict the number of mutations that are required for a tumour to evolve[5]. Doll developed a model in which the time taken for cancer to arise depends on the probabilities of each of the mutations needed to cause the cancer actually occurring, and he fitted the model to real incidence statistics. But Michor says that this approach fails to take into account population-genetics theory. Doll's models look at single cells, ignoring the fact that if the first mutation increases the evolutionary fitness of that cell, then the mutation will expand into many cells, increasing the probability that subsequent mutations will occur.

Understanding this population effect will be hugely important in overcoming evolved drug resistance. One way to deal with this is to use combinations of treatments that tackle different aspects of the disease. "We can try to come up with treatment strategies if we understand how many mutations are needed for resistance," says Michor. "We can actually write down the equations that predict what the risk of resistance is depending on how many drugs you use." A crucial goal will be to make these models predict what is going on in systems with complicated evolutionary trajectories, such as AML.

"Cancer genomes help us get quantitative data to calibrate our evolutionary models."
— Martin Nowak

Pushing the parallels

So how far can the evolutionary idea be extended? The small number of cell divisions within a cancer compared with that in the evolution of species is an obvious limitation. But there are still plenty of evolutionary ideas to be explored in cancer, several of which come from thinking about the whole ecosystem of the disease. As with 'real' ecosystems, these involve not just the species in question, but also its competitors, predators and symbionts.

The evolving cancer cell not only needs to outcompete the normal body cells, it must also evade attack by the immune system and, if it is to reach the advanced stages of cancer, it needs to cooperate with other cells and then migrate and colonize other parts of the body. For example, once a tumour reaches a certain critical mass its cells require a blood supply to keep them oxygenated. This means that it needs to co-opt the body's system for creating blood vessels.

How these processes take place is ripe ground for evolutionary biologists and ecologists to investigate, something that Maley and his colleague John Pepper, of the University of Arizona in Tucson, were keen to encourage when they organized a recent workshop on the topic at the Santa Fe Institute in New Mexico. The fact that a similar workshop was organized by the National Cancer Institute, a major funding body, suggests that the money might follow. Maley certainly hopes so. "I see my role," he says, "as attempting to bring evolutionary biologists into cancer biology and advocating the need for evolutionary biologists as part of our interdisciplinary teams." Whether such interdisciplinary research will entice evolutionary biologists to shift their field study to the clinic is yet to be seen, but for investigations of variation and selection, cancers unfortunately continue to produce ample material for study. ■

Patrick Goymer is associate editor of *Nature Reviews Genetics*.

1. Campbell, P. J. *et al. Blood* **108,** 3548–3555 (2006).
2. Nowell, P. C. *Science* **194,** 23–28 (1976).
3. Maley, C. C. *et al. Nature Genet.* **38,** 468–473 (2006).
4. Jones, S. *et al. Proc. Natl Acad. Sci. USA* **105,** 4283–4288 (2008).
5. Armitage, P. & Doll, R. *Br. J. Cancer* **8,** 1–12 (1954).
6. Varshavsky, A. *Proc. Natl Acad. Sci. USA* **104,** 14935–14940 (2007).

CORRESPONDENCE

Postdoc glut means academic pathway needs an overhaul

SIR — In his Correspondence 'Fewer academics are not the answer to funding woes' (*Nature* **454,** 397; 2008), Philip Strange suggests that we need to increase the number of trained scientists to help deal with current and future crises such as climate change. But, according to the US National Science Board's Science and Engineering Indicators 2008 (www.nsf.gov/statistics/seind08), fewer than 20% of postdoctoral scientists in the United States find tenure-track faculty positions. This suggests that, at least in the United States, we could already have a glut of trained scientists.

Perhaps the solution is not financial at its core at all. A major overhaul of the academic training pathway for life-scientists is long overdue. Issues linked to today's financial and job markets are an indicator that the time is right for a serious self-appraisal on the part of academia. Are we training too many students? And what should we do with all the postdocs?

Ian M. Brooks UTHSC Postdoctoral Association, and Department of Pharmacology, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA
e-mail: ibrooks1@utmem.edu

Olympics may have a negative impact on China's research

SIR — Your Editorial 'China's challenges' (*Nature* **454,** 367–368; 2008) and News Feature 'Visions of China' (*Nature* **454,** 384–387; 2008) proclaim China's increasing strength in science and technology. But this summer's Olympics Games in Beijing may cast a shadow over that landscape.

The Chinese government's investment in the 2008 Beijing Olympics is reflected in a wide range of technical accomplishments. These include the design and construction of the venues, development of energy-saving technology, communication support, comprehensive services for the media, public information and weather forecasting, advanced safety and security, and rigorous drug testing.

But away from the extravagant displays of the opening ceremony, scientists in China are experiencing some negative effects that could take hold in the wake of the Olympics. Nationwide security restrictions on the transport of liquids, for example, have delayed the delivery of chemical and biological reagents to laboratories, especially those coming from abroad. The resources of the entire country have been tapped, draining funds from fields outside Olympic interests. Add to this the cost of the recent earthquakes in Sichuan, and it will become even harder to meet the Chinese state's optimistic goals for research and development.

The post-Olympic era is likely to see more emphasis on disciplines such as Earth science and environmental protection, but considerably less on other areas of science and technology.

Yijun Chen, Nan Liu Laboratory of Chemical Biology, China Pharmaceutical University, Nanjing, Jiangsu 210009, China
e-mail: yjchen@cpu.edu.cn

Changes in the rules now governing Italy's drug industry

SIR — I was surprised and dismayed that your Editorial 'Clean hands, please' (*Nature* **454,** 667; 2008) should be so careless with the evidence in addressing the question of relations between the government and the pharmaceutical industry in Italy.

In Italy, the pharmaceutical industry is obliged to maintain the lowest prices in Europe. The new government has left these regulations unaltered in its economic programme for the next three years.

At the time of the unfortunate events of the Duilio Poggiolini era 15 years ago, Italy's per capita spending on pharmaceuticals was slightly below the European average. This spending has today fallen to €194 (US$290), against a European average of €270 — even though Italy has easier access to therapies than many other European countries and has more elderly citizens than any of them.

In the past seven years, governments of various political colours have made 18 separate cutbacks on pharmaceutical spending. The 30% difference from the European average indicates that our industry has not enjoyed government favours.

The pharmaceutical industry has nevertheless fought back against this difficult situation. Its exports have risen from 10% of total production to 53%, and today it boasts a surplus in its trade balance for medicines.

The replacement of the manager of the Italian medicines agency AIFA, as mentioned in your Editorial, occurred after a judicial enquiry. This was ratified by an independent judicial decision.

The pharmaceutical industry in Italy is represented by an association that belongs to the Italian employers' federation Confindustria. This association has implemented rigorous ethical rules to govern its members' conduct and, in representing the interests of the industry *vis-à-vis* the public institutions, totally respects their autonomy. Its sole request is that public institutions provide a regulatory framework that is fixed over time and can guarantee timescales and procedures in their authorization processes, in the same way as those in force in other European countries.

Sergio Dompé Farmindustria, Largo del Nazareno 3, 00187 Roma, Italy
e-mail: morelli@farmindustria.it

Atheism could be science's contribution to religion

SIR — We were perplexed by your Editorial on the work of the Templeton Foundation ('Templeton's legacy' *Nature* **454,** 253–254; 2008). Surely science is about finding material explanations of the world — explanations that can inspire those spooky feelings of awe, wonder and reverence in the hyper-evolved human brain.

Religion, on the other hand, is about humans thinking that awe, wonder and reverence are the clue to understanding a God-built Universe. (The same is true of religion's poor cousin, 'spirituality', which you slip into your Editorial rather as a creationist uses 'intelligent design'.) There is a fundamental conflict here, one that can never be reconciled until all religions cease making claims about the nature of reality.

The scientific study of religion is indeed full of big questions that need to be addressed, such as why belief in religion is negatively correlated with an acceptance of evolution. One could consider psychological studies of why humans are superstitious and believe impossible things, and comparative sociological studies of religion using materialist explanations of the rise and fall of the world's belief systems.

Perhaps the Templeton Foundation is thinking of funding such research. The outcome of such work, we predict, will not bring science and religion (or 'spirituality') any closer to one another. You suggest that science may bring about "advances in theological thinking". In reality, the only contribution that science can make to the ideas of religion is atheism.

Matthew Cobb Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, UK
e-mail: cobb@manchester.ac.uk
Jerry Coyne Department of Ecology and Evolution, The University of Chicago, Chicago, Illinois 60637, USA

BOOKS & ARTS

A. SHINBUN

The blossoming of Japanese mathematics

A new compilation of the illustrated geometry problems that decorated shrines in seventeenth-century Japan provides puzzles that are still intriguing today, finds Peter J. Lu.

Sacred Mathematics: Japanese Temple Geometry
by Fukagawa Hidetoshi and Tony Rothman
Princeton University Press: 2008. 392 pp.
$35.00, £19.95

At the beginning of the seventeenth century, Tokugawa Ieyasu completed the unification of Japan. His shogunate ruled for more than 250 years and oversaw a period of peace, but with restricted foreign contact. Poetry, music and literature flourished during this time of relative isolation. A unique form of Japanese culture of the period was *sangaku* — a combination of mathematics and art on votive tablets. Illustrated wooden shingles up to several metres across bearing geometry problems were hung from Shinto shrines and Buddhist temples for public display. Many historical *sangaku* answers appeared on tablets without proof, perhaps to demonstrate the mathematical prowess of the presenter. And their sacred context remains unclear: the tablets might have been educational or may have signalled gratitude for divine assistance in solving a mathematical problem. Of the thousands of tablets created, only a fraction survive, and they have received scant coverage in histories of Japanese mathematics.

Now Fukagawa Hidetoshi, a mathematics teacher, and writer Tony Rothman present a collection of *sangaku* problems in their book, *Sacred Mathematics*. The puzzles range from simple algebra within the grasp of any intermediate-school student, to challenging problems that require graduate-school mathematics to solve. Copious illustrations and many detailed solutions show the scope, complexity and beauty of what was tackled in Japan during the Tokugawa shogunate.

The book offers a feast for recreational geometers looking for fun new problems, presented and solved in clever ways. Yet the authors give little insight into how these problems were solved at the time, or whether unique Japanese methods were involved. The *sangaku* figures are mostly redrawn with modern notation, and solutions offered in the compact form of present-day Western mathematics. Illustrating traditional Japanese and modern Western methods side by side would have been instructive. The book thus achieves only limited success in showcasing *sangaku* as exemplars of a uniquely Japanese style of mathematics, because that style is never elucidated.

Fukagawa and Rothman illuminate the mathematics more than the history and context of the tablets. Citations to mathematical theorems abound, yet references supporting their historical claims are absent. More seriously, the historical commentary reflects a romantic bias that a unique Japanese culture flowered because of its complete isolation. The book states, for instance, that "a unique brand of homegrown mathematics flourished, one that was completely uninfluenced by developments in the western mathematics". This generalization is historically unsupported, and obstructs an accurate consideration of the interplay of factors that drove the development of Japanese mathematics.

Along similar lines, the authors also dismiss the millennium that preceded the seventeenth century as "a dark age" paralleling that in Europe, during which relatively little was accomplished in mathematics. By ignoring the medieval Islamic world, they fall into the same trap as Eurocentric mathematical historians who focus exclusively on ancient Greece and modern Europe. This omission undermines their discussion of the Chinese foundation of Japanese mathematics. World-leading achievements in mathematics, science and technology — even astronomers from Persia — reached China during the medieval period by the Silk Road and other routes. The book describes in great detail how Seki Takakazu, "Japan's most celebrated mathematician", calculated π to 11 digits in the eighteenth century. But it does not mention Jamshid Masʿud al-Kashi, who determined π correctly to 16 digits some three centuries earlier while residing in Samarkand (in what is now Uzbekistan), one of the most important cities along the Silk Road.

The complex events that followed Japan's opening to the West after the shogunate's end in the 1860s are given similarly short shrift. The response of Japanese mathematics to this influx of ideas is dispatched with the glib statement that "resistance was futile". By quoting a science-fiction character from *Star Trek* that annihilates everything in its path by assimilation, Fukagawa and Rothman trivialize Japan's complicated process of reintegration with the international community, and miss the opportunity to shed light on how its early mathematics contributed to Japan's present-day leadership in science and technology. ■

Peter J. Lu is a postdoctoral research fellow in the department of physics at Harvard University, Cambridge, Massachusetts 02138, USA.
e-mail: plu@fas.harvard.edu

The future ain't what it used to be

Future Proof/You Call This the Future?
by Nick Sagan
with Mark Frary and Andy Walker
Icon Books/Chicago Review Press: 2008. 160 pp. £10.99/$14.95

Rather than teleporting, or even arriving on a hoverboard, I rode to work this morning on the distinctly nineteenth-century technology of a bicycle on tarmac. Much to the relief of my colleagues, I don't sit at my desk in a silver jumpsuit, and my lunch doesn't come in pill form. As a confirmed sci-fi geek, I am perpetually disappointed that the things promised to me by science fiction have not yet arrived.

Good-quality sci-fi is concerned with the present, our current fears and hopes. The technology imagined either drives plot, such as the human-powered virtual reality of the *Matrix* movies, or is set-dressing, as in the film *Alien* — at its core a haunted-house story. Alas, accurate fortune-telling in science fiction is rare. No one really predicted the dominance of the personal computer over our lives, nor that the rather low-tech medium of text-messaging would become such a popular form of communication. As baseball legend Yogi Berra put it: "It's tough to make predictions, especially about the future."

And that is the subject of Nick Sagan's terrific new book. Sagan, a Hollywood screenwriter, carries a healthy pedigree in this field. He's the son of Carl Sagan who, aside from his work as a cosmologist, also wrote some outstanding sci-fi, including *Contact*. The excellent film version with Jodie Foster in the lead is one of the less fanciful sci-fi films of the 1990s.

Future Proof is the latest in a recent run of guidebooks, such as Daniel H. Wilson's *Where's My Jetpack?*, about the fictional products imagined by sci-fi's finest creators and how close we are to realizing those phenomena. Sagan's journey includes terraforming by cyborgs and teleporting through wormholes. This snapshot of all the bleeding-edge technology currently available, even though it falls short of the fiction, suggests that human endeavour is a truly glorious thing — one that we too often take for granted. With his co-authors Mark Frary and Andy Walker, Sagan writes casually well. And the book features handy diagrams and references to all the best twentieth-century sci-fi.

UNIV. WASHINGTON

Lenses that can project in-eye displays now exist.

The only trouble is that, by definition, science moves so quickly that some entries in the book are already out of date. For example, Babak Parviz and his colleagues from the University of Washington in Seattle recently announced that they are close to completing a contact lens with embedded light-emitting diodes, the first step towards creating an in-eye display reminiscent of Arnold Schwarzenegger's relentless Terminator. Meanwhile, researchers from Imperial College London have published a paper describing an invisibility cloak — albeit still theoretical, and highly impractical.

Nevertheless, *Future Proof* is a fun, useful primer for understanding the symbiosis between science and science fiction. Perversely, I hope that this version of the book becomes redundant sooner rather than later.

As an experiment in testing the limitations of a highly futuristic invention, I wrote this review on my iPhone. It's a touch-screen device with so many slick functions that it sometimes feels indistinguishable from magic: when I first turned it on, it creepily showed me where I was standing. I've become accustomed to its ways so quickly that I often find myself smudging non-touch-screens in error. But as a typewriter for a 500-word book review, it was hopeless. I look forward to the day when my children will puzzle and sneer at this crude silicon-based tricorder prototype. Until then, I'm not imaginative or foolish enough to dream up what their essential devices might be. ■

Adam Rutherford is Nature Publishing Group's Podcast and Video editor.

More cacophony than harmony

The World In Six Songs: How the Musical Brain Created Human Nature
by Daniel J. Levitin
Dutton Books: 2008. 333 pp. $25.95

Six songs seems a small repertoire to address so grandiose a theme. Yet Daniel Levitin contentiously argues for six classes of song in his quickly published follow-on from *This Is Your Brain On Music*. He purports to explain neuroscientific concepts by framing them around song themes of friendship, joy, comfort, knowledge, religion and love.

Notwithstanding his claim that he "endeavoured to include examples from music [from] all over the world", Levitin, a recording producer turned academic psychologist, remains tenacious in his devotion to pop music. Granted, he alludes to a love for Beethoven's *Pastoral Symphony* and mentions Mozart occasionally, but other musical genres seem of far less importance to him. Listing "the musical events that changed the way I would hear for the rest of my life", every one is popular: Sting, Cannonball Adderley and Paul Simon are among his favourites. It would be unthinkable for the author of a book on, say, the 'pictorial brain' or the 'literary brain' to ignore the greatest examples of those arts or to reveal so enthusiastically that his tastes seem not to have matured since his adolescence.

Levitin writes for a US audience of a certain age that has never outgrown its attachment to the music of its youth. There is little in *The World in Six Songs* to acknowledge those with matured or non-American musical tastes, which creates difficulties when he writes about "us" having "common ground" or indulges in such hyperbole as "our truth detectors go wild". Who, apart from US baby-boomers, might comprise this "us"?

Far from restricting himself to six songs, and setting aside that a different person might classify those songs differently, the author's repertoire of musical examples is vast. His collection of 'friends' seems equally extensive. The text is littered with names — musicians and research academics among them — and their mention often seems calculated to impress rather than enlighten. Every significant thinker in his field sounds like a trusted intimate.

This makes Levitin look fortunate but, even if it were true, makes him intellectually vulnerable. One is unlikely to be critical of the work of such 'friends'. And so it proves. When

The future ain't what it used to be

Future Proof/You Call This the Future?
by Nick Sagan
with Mark Frary and Andy Walker
Icon Books/Chicago Review Press: 2008. 160 pp. £10.99/$14.95

Rather than teleporting, or even arriving on a hoverboard, I rode to work this morning on the distinctly nineteenth-century technology of a bicycle on tarmac. Much to the relief of my colleagues, I don't sit at my desk in a silver jumpsuit, and my lunch doesn't come in pill form. As a confirmed sci-fi geek, I am perpetually disappointed that the things promised to me by science fiction have not yet arrived.

Good-quality sci-fi is concerned with the present, our current fears and hopes. The technology imagined either drives plot, such as the human-powered virtual reality of the *Matrix* movies, or is set-dressing, as in the film *Alien* — at its core a haunted-house story. Alas, accurate fortune-telling in science fiction is rare. No one really predicted the dominance of the personal computer over our lives, nor that the rather low-tech medium of text-messaging would become such a popular form of communication. As baseball legend Yogi Berra put it: "It's tough to make predictions, especially about the future."

And that is the subject of Nick Sagan's terrific new book. Sagan, a Hollywood screenwriter, carries a healthy pedigree in this field. He's the son of Carl Sagan who, aside from his work as a cosmologist, also wrote some outstanding sci-fi, including *Contact*. The excellent film version with Jodie Foster in the lead is one of the less fanciful sci-fi films of the 1990s.

Future Proof is the latest in a recent run of guidebooks, such as Daniel H. Wilson's *Where's My Jetpack?*, about the fictional products imagined by sci-fi's finest creators and how close we are to realizing those phenomena. Sagan's journey includes terraforming by cyborgs and teleporting through wormholes. This snapshot of all the bleeding-edge technology currently available, even though it falls short of the fiction, suggests that human endeavour is a truly glorious thing — one that we too often take for granted. With his co-authors Mark Frary and Andy Walker, Sagan writes casually well. And the book features handy diagrams and references to all the best twentieth-century sci-fi.

UNIV. WASHINGTON

Lenses that can project in-eye displays now exist.

The only trouble is that, by definition, science moves so quickly that some entries in the book are already out of date. For example, Babak Parviz and his colleagues from the University of Washington in Seattle recently announced that they are close to completing a contact lens with embedded light-emitting diodes, the first step towards creating an in-eye display reminiscent of Arnold Schwarzenegger's relentless Terminator. Meanwhile, researchers from Imperial College London have published a paper describing an invisibility cloak — albeit still theoretical, and highly impractical.

Nevertheless, *Future Proof* is a fun, useful primer for understanding the symbiosis between science and science fiction. Perversely, I hope that this version of the book becomes redundant sooner rather than later.

As an experiment in testing the limitations of a highly futuristic invention, I wrote this review on my iPhone. It's a touch-screen device with so many slick functions that it sometimes feels indistinguishable from magic: when I first turned it on, it creepily showed me where I was standing. I've become accustomed to its ways so quickly that I often find myself smudging non-touch-screens in error. But as a typewriter for a 500-word book review, it was hopeless. I look forward to the day when my children will puzzle and sneer at this crude silicon-based tricorder prototype. Until then, I'm not imaginative or foolish enough to dream up what their essential devices might be. ■

Adam Rutherford is Nature Publishing Group's Podcast and Video editor.

More cacophony than harmony

The World In Six Songs: How the Musical Brain Created Human Nature
by Daniel J. Levitin
Dutton Books: 2008. 333 pp. $25.95

Six songs seems a small repertoire to address so grandiose a theme. Yet Daniel Levitin contentiously argues for six classes of song in his quickly published follow-on from *This Is Your Brain On Music*. He purports to explain neuroscientific concepts by framing them around song themes of friendship, joy, comfort, knowledge, religion and love.

Notwithstanding his claim that he "endeavoured to include examples from music [from] all over the world", Levitin, a recording producer turned academic psychologist, remains tenacious in his devotion to pop music. Granted, he alludes to a love for Beethoven's *Pastoral Symphony* and mentions Mozart occasionally, but other musical genres seem of far less importance to him. Listing "the musical events that changed the way I would hear for the rest of my life", every one is popular: Sting, Cannonball Adderley and Paul Simon are among his favourites. It would be unthinkable for the author of a book on, say, the 'pictorial brain' or the 'literary brain' to ignore the greatest examples of those arts or to reveal so enthusiastically that his tastes seem not to have matured since his adolescence.

Levitin writes for a US audience of a certain age that has never outgrown its attachment to the music of its youth. There is little in *The World in Six Songs* to acknowledge those with matured or non-American musical tastes, which creates difficulties when he writes about "us" having "common ground" or indulges in such hyperbole as "our truth detectors go wild". Who, apart from US baby-boomers, might comprise this "us"?

Far from restricting himself to six songs, and setting aside that a different person might classify those songs differently, the author's repertoire of musical examples is vast. His collection of 'friends' seems equally extensive. The text is littered with names — musicians and research academics among them — and their mention often seems calculated to impress rather than enlighten. Every significant thinker in his field sounds like a trusted intimate.

This makes Levitin look fortunate but, even if it were true, makes him intellectually vulnerable. One is unlikely to be critical of the work of such 'friends'. And so it proves. When

the breathless pace of his prose leads him to make glib generalizations, a check of a few sources creates serious doubts about the way in which he summarizes their research findings. Perhaps he has not had time to assimilate them adequately. Levitin's text gives the impression of an assignment by a highly intelligent student who has done so much preliminary reading and assembled so many references that he cannot discipline himself to omit a single one. Not all research is good or worthwhile and, in such a book, more is not necessarily better. Stricter selectivity would have helped.

The writing is a problem, too. Much of it reads as if it was dictated into a machine and given the most cursory of edits. Many sentences are needlessly opaque. Some paragraphs are trivial and others are contradictory; his accounts of outcomes of experiments differ — a serious flaw — and incompatible 'facts' are given, such as about the sensitivity of auditory hair cells. In places the writing seems lazy, as with his over-use of the phrase "neurochemical soup". One questions his understanding of some material, such as an assertion that "grammar specifies rules … based on where in the sentence the elements appear", which betrays an unawareness of Latin and German. Worst of all are the pages of self-indulgent digression from his grand, and essentially unprovable, thesis that "the musical brain created human nature". The book should have been much shorter and better constructed.

Why has Levitin written it? I suspect that it was at the publisher's urging, after the success of his far superior earlier book. The reader's time would be better spent with that previous book or by engaging with the research articles that Levitin draws on. One would then also be spared the boastful confusion of friends and anecdotes, the sheer volume of which the scientific brain, let alone the musical brain, struggles to deal with. ■

John Carmody is a medical scientist at the University of Sydney, Australia, who writes and broadcasts on music and scientific history.
e-mail: jcarmody@med.usyd.edu.au

Innovations of an ancient nation

Chinese Memory: Treasures of a 5,000-year-old Civilization
Capital Museum, Beijing
Until 7 October 2008

This ancient Chinese 'environmentally friendly' oil lamp (pictured) has a built-in system to eliminate smoke and dates from the Western Han dynasty of 206 BC to AD 25. It is one of 169 artefacts on display in the exhibition *Chinese Memory*, showing at the Capital Museum in Beijing until October. The items, ranging from a 6,000 BC flute to a twentieth-century incense burner, were chosen from 55 museums across the country to trace China's technological and cultural innovations.

The bronze oil lamp was excavated in Pingsu, Shanxi province, in 1985. It takes the shape of a swan goose turning its neck backwards, holding a fish in its mouth. A round burner sits on the bird's back, shaded by the fish. Two curved bronze panels make an adjustable shutter for the burner, and can be slid open or closed to shelter the flame from wind and to control the illumination level. When the lamp is lit, smoke rises into the bird's mouth and is channelled via the neck into the belly, where it is trapped by water. Assembled from four components, it can be taken apart for cleaning.

The lamp's form reflects a fusion of philosophies. During the early Han period, the emperors co-opted many ideologies that had been suppressed by the preceding Qin dynasty (221–206 BC), during which scholarly books were burned and Confucians buried alive. The lamp combines Confucianism's regard for utility with Taoism's respect for living in accord with nature. The swan goose symbolizes compassion and blessing and the fish represents prosperity.

By embodying the traditional Chinese ideal of *tianren heyi* — the harmony between man and nature — this exquisite lamp's design points to the early desire among the ancients to live in a clean environment. ■

Jane Qiu is a science writer based in London and Beijing.
e-mail: jane@janeqiu.com

CULTURE DISH

ART OF NOISE
An early collaboration between engineers at Bell Laboratories in New Jersey and ten artists, among them Robert Rauschenberg, Robert Whitman and John Cage, is highlighted in an exhibition at New York's Museum of Modern Art. The group built and featured novel technical equipment in avant-garde theatre and dance performances in October 1966 in New York. This formed part of a wider programme by Experiments in Art and Technology, an organization that promoted crossover projects until the 1980s.


IMAGE COURTESY OF D. BOWIE

Looking at Music, which runs until January 2009, explores how music, art and technology have influenced each other since the innovations of video and electric guitars in the 1960s. On display are video, audio and photographic works by artists such as Laurie Anderson, Nam June Paik and Bruce Nauman. Films accompany the show and track developments in technologies, including those in early music videos by The Beatles, David Bowie (pictured) and Captain Beefheart
http://tinyurl.com/5gakt6

SMILES OF SUCCESS


P. BADGE/WILEY-BLACKWELL

Portraits of Nobel prizewinners by the German photographer Peter Badge are collected in *Nobels*, published by Wiley this month. Badge has tracked down and photographed more than 300 living Nobel laureates for this glossy coffee-table book, including economist John Nash and physicist Hans Dehmelt (pictured). A biography is presented alongside each portrait.
http://tinyurl.com/5qmu7s

CULTURE DISH

ART OF NOISE

An early collaboration between engineers at Bell Laboratories in New Jersey and ten artists, among them Robert Rauschenberg, Robert Whitman and John Cage, is highlighted in an exhibition at New York's Museum of Modern Art. The group built and featured novel technical equipment in avant-garde theatre and dance performances in October 1966 in New York. This formed part of a wider programme by Experiments in Art and Technology, an organization that promoted crossover projects until the 1980s.

IMAGE COURTESY OF D. BOWIE

Looking at Music, which runs until January 2009, explores how music, art and technology have influenced each other since the innovations of video and electric guitars in the 1960s. On display are video, audio and photographic works by artists such as Laurie Anderson, Nam June Paik and Bruce Nauman. Films accompany the show and track developments in technologies, including those in early music videos by The Beatles, David Bowie (pictured) and Captain Beefheart
http://tinyurl.com/5gakt6

SMILES OF SUCCESS

P. BADGE/WILEY-BLACKWELL

Portraits of Nobel prizewinners by the German photographer Peter Badge are collected in *Nobels*, published by Wiley this month. Badge has tracked down and photographed more than 300 living Nobel laureates for this glossy coffee-table book, including economist John Nash and physicist Hans Dehmelt (pictured). A biography is presented alongside each portrait.
http://tinyurl.com/5qmu7s

the breathless pace of his prose leads him to make glib generalizations, a check of a few sources creates serious doubts about the way in which he summarizes their research findings. Perhaps he has not had time to assimilate them adequately. Levitin's text gives the impression of an assignment by a highly intelligent student who has done so much preliminary reading and assembled so many references that he cannot discipline himself to omit a single one. Not all research is good or worthwhile and, in such a book, more is not necessarily better. Stricter selectivity would have helped.

The writing is a problem, too. Much of it reads as if it was dictated into a machine and given the most cursory of edits. Many sentences are needlessly opaque. Some paragraphs are trivial and others are contradictory; his accounts of outcomes of experiments differ — a serious flaw — and incompatible 'facts' are given, such as about the sensitivity of auditory hair cells. In places the writing seems lazy, as with his over-use of the phrase "neurochemical soup". One questions his understanding of some material, such as an assertion that "grammar specifies rules … based on where in the sentence the elements appear", which betrays an unawareness of Latin and German. Worst of all are the pages of self-indulgent digression from his grand, and essentially unprovable, thesis that "the musical brain created human nature". The book should have been much shorter and better constructed.

Why has Levitin written it? I suspect that it was at the publisher's urging, after the success of his far superior earlier book. The reader's time would be better spent with that previous book or by engaging with the research articles that Levitin draws on. One would then also be spared the boastful confusion of friends and anecdotes, the sheer volume of which the scientific brain, let alone the musical brain, struggles to deal with. ■

John Carmody is a medical scientist at the University of Sydney, Australia, who writes and broadcasts on music and scientific history.
e-mail: jcarmody@med.usyd.edu.au

Innovations of an ancient nation

Chinese Memory: Treasures of a 5,000-year-old Civilization
Capital Museum, Beijing
Until 7 October 2008

This ancient Chinese 'environmentally friendly' oil lamp (pictured) has a built-in system to eliminate smoke and dates from the Western Han dynasty of 206 BC to AD 25. It is one of 169 artefacts on display in the exhibition *Chinese Memory*, showing at the Capital Museum in Beijing until October. The items, ranging from a 6,000 BC flute to a twentieth-century incense burner, were chosen from 55 museums across the country to trace China's technological and cultural innovations.

The bronze oil lamp was excavated in Pingsu, Shanxi province, in 1985. It takes the shape of a swan goose turning its neck backwards, holding a fish in its mouth. A round burner sits on the bird's back, shaded by the fish. Two curved bronze panels make an adjustable shutter for the burner, and can be slid open or closed to shelter the flame from wind and to control the illumination level. When the lamp is lit, smoke rises into the bird's mouth and is channelled via the neck into the belly, where it is trapped by water. Assembled from four components, it can be taken apart for cleaning.

The lamp's form reflects a fusion of philosophies. During the early Han period, the emperors co-opted many ideologies that had been suppressed by the preceding Qin dynasty (221–206 BC), during which scholarly books were burned and Confucians buried alive. The lamp combines Confucianism's regard for utility with Taoism's respect for living in accord with nature. The swan goose symbolizes compassion and blessing and the fish represents prosperity.

By embodying the traditional Chinese ideal of *tianren heyi* — the harmony between man and nature — this exquisite lamp's design points to the early desire among the ancients to live in a clean environment. ■

Jane Qiu is a science writer based in London and Beijing.
e-mail: jane@janeqiu.com

Spartan sport laid bare

Edgar Degas's painting of female athletes challenging male competitors in classical Sparta raises subtle questions about gender, politics and sport, explains Martin Kemp.

J. WEBB/NATL GALLERY, LONDON/THE ART ARCHIVE

Degas's *The Young Spartans* portrays an ancient scene yet highlights modern ideas about feminism.

The original Olympic Games, first recorded in 776 BC at Olympia, Greece, were driven by the Greek cult of the body beautiful and the body athletic. This is directly reflected in ancient sculptures of the male form. Revered physiques were of two types — the slender and graceful form typified by Apollo, and the rugged muscularity of Hercules.

Among early Greek cultures, the militaristic state of Sparta alone accorded women a status comparable with that of men in the cultivation of an athletic physique. Plutarch tells us that Lycurgus, the severe law-giver of Sparta, "ordered the maidens to exercise themselves with wrestling, running, throwing the quoit, and casting the dart, to the end that the fruit they conceived might, in strong and healthy bodies, take firmer root and find better growth".

This ideal was accompanied by the routine exposure of naked or partially clothed women's bodies, which was seen as modest rather than shameful, with all wantonness excluded. Plutarch says that this "taught them simplicity and a care for good health, and gave them some taste of higher feelings, admitted as they thus were to the field of noble action and glory".

All this contrasts with the better known Athenian culture, in which women were largely subordinated in domestic roles and were certainly not expected to display their flesh. Athenian standards set the tone for subsequent European societies. Yet the Spartan cult of bodily resolution, based on austere physical training and forbearance of pain, has inspired many military and political movements, including the French Revolution. The word spartan, signifying courageous endurance, has entered the modern vocabulary.

Sparta has also bequeathed to us the adjective 'laconic', after the name of the territory Laconia, over which the city of Sparta ruled. The Spartans were renowned for being people of few words, preferring brief utterances and decisive action to literary and rhetorical flourishes.

The singular position of women in Spartan culture attracted the attention of the French artist Hilaire-Germain-Edgar Degas around 1860. Those who know Degas from his famous scenes of contemporary life, especially his bathers, dancers and horse races, may be surprised to find him tackling the kind of ancient subject that was favoured by the conservative academy. Degas, however, made a very modern subject from ancient history.

The painting *The Young Spartans* was begun around 1860 and remained in Degas's studio. He planned to exhibit it in partly reworked form at the Fifth Impressionist Exhibition in 1880, but did not do so. His patron and friend, Diego Martelli, described the subject as "girls who challenge the boys to a race, which decided, in accordance with the law of those people, their submission".

On an open grassy plain, four lightly clad young women confront a group of naked young men, who have cast off their rudimentary garments. The youths posture self-consciously as adolescent males will, exercising a precocious bravado. The leading girl issues a challenge or provocation. In the background, as Degas has testified, is the white-robed Lycurgus with a group of mothers and their children.

Compared with a prior, unfinished version of the painting in the Art Institute of Chicago in Illinois, the antique elements are played down. A classical shrine has been eliminated, and the foreground protagonists have lost their ancient appearance. They are now cast as the kind of lower class Parisian youths whom Degas sometimes sketched.

Degas was deeply interested in physiognomy, ethnography, anthropology and criminology. His famous sculpture of the *Little Dancer of Fourteen Years* was exhibited in a glass case in 1881 as a 'primitive' ethnographic type, alongside his pastels of criminal physiognomies. His Spartan boys exhibit the kind of profiles with protruding jaw lines that, in a post-Darwinian world, were seen as regressive. They have distinctly animalistic traits, not least in the case of the blond youth on all fours.

The exceptional characterization of the women as being openly able to challenge their prospective mates for supremacy may well be linked to the French feminist ideas of around 1880. Martelli was an outspoken liberal and supporter of women's causes, not least those of prostitutes, whom he saw as brutalized by the sexual desires of men. 'Brutalized' is itself a loaded word in this post-Darwinian context.

Degas's painting is therefore not a simple homage to classical antiquity in the traditional manner of academic history paintings. It stands in a complex relationship with what was known about a remarkable early society and its potential significance for contemporary issues. Politics, feminism and anthropology are all in the artistic mix.

It is worth noting that the first of Baron Pierre de Coubertin's modern Olympic Games, held in Athens in 1896, included only male athletes. Women competitors were admitted four years later to the second Olympics in Paris, that most modern and political of contemporary cities. Yet unlike Sparta, the women did not compete directly with the men. ■

Martin Kemp is research professor in the history of art at the University of Oxford, OX1 1PT, UK, and author of *The Human Animal in Western Art and Science*.

ESSAY

Battle of the sexes may set the brain

A tug-of-war between the mother's and father's genes in the developing brain could explain a spectrum of mental disorders from autism to schizophrenia, suggest **Christopher Badcock** and **Bernard Crespi**.

It has long been recognized that mental illnesses such as schizophrenia and autism tend to run in families. But neither disorder obeys classical Mendelian laws of inheritance, making it difficult to pinpoint the genes involved.

We believe that psychiatric illness may be less to do with the genes a mother and father pass down, and more to do with which genes they program for expression. By our hypothesis, a hidden battle of the sexes — where a mother's egg and a father's sperm engage in an evolutionary struggle to turn gene expression up or down — could play a crucial part in determining the balance or imbalance of an offspring's brain. If this proves true, it would greatly clarify the diagnosis of mental disorders. It might even make it possible to reset the mind's balance with targeted drugs.

The story begins with the late William Hamilton, originator of the selfish gene idea popularized by Richard Dawkins. Hamilton explained how genetic traits that are apparently injurious can evolve by natural selection. He proved that genes predisposing an individual to self-sacrifice could propagate in a population if they were sufficiently shared by the beneficiaries of such acts. Charles Darwin did not know about genes, but saw evolution in terms of conflict between individuals, groups and species. Hamilton showed that evolution is, in fact, a question of conflict between genes. Because individuals in sexually reproducing species receive genes from two different parents, he also realized that such conflicts probably occur within the same individual.

Shortly before Hamilton's untimely death in 2000, evidence was found for this internal selfish-gene behaviour. Geneticists discovered that some crucial genes are expressed when inherited from one parent, but not when inherited from the other. This is achieved through a process called imprinting, in which genes in the sperm and egg are marked for expression or silencing in a later embryo and child. For example, a fetus inherits a gene called *IGF2*, which encodes an insulin-like growth factor, from both its mother and its father. In mammals, only the father's copy is normally expressed. In humans, if the mother's copy is also expressed, the result is a child with Beckwith–Wiedemann syndrome. This is characterized by a birthweight that is more than 50% above normal, with various other symptoms of over-growth. If, by contrast, both the father's and the mother's copies are silenced, the opposite outcome occurs — under-growth, as featured in Silver–Russell syndrome.

Larger babies live longer, develop less disease and have better all-round health. It is now thought that the genetic conflict over offspring size arises because the father's genes gain these benefits by being carried in a large offspring at no personal cost to the father. The mother, however, pays the costs of gestating, giving birth to and suckling a larger child. Hence, mammalian mothers silence their copies of growth-enhancing genes such as *IGF2*, whereas fathers mark them for expression.

Genes that are either maternally or paternally biased thus engage in a genetic tug-of-war. Letting the rope slip in one direction or the other leads to opposite outcomes in offspring.

Powerful pull

It is currently thought that perhaps a few hundred of the 20,000 or so human genes are subject to imprinting, although only 63 are confirmed so far. This number might sound small, but imprinted genes such as *IGF2* commonly have far-reaching effects on growth and development. Imprinting has been found to be especially frequent in genes expressed in the placenta, the organ that governs how resources are extracted from the mother. Imprinting is also common among genes that drive brain development. A reasonable conclusion is that genetic tugs-of-war should also affect behaviour, cognition and personality.

> "To propose a single, overriding explanation for such a huge range of mental conditions is controversial."

Hamilton certainly thought that genetic conflict would have psychological consequences. He noticed that there were 'people people' and 'things people', classifying himself as the latter. Although these tendencies could be "disastrous socially", he wrote, "I believe it is in essence an aberration of this kind that makes me a successful scientist"[1]. When taken to an extreme, the social difficulties and mechanistic tendencies of 'things people' are recognizable as symptoms of autism.

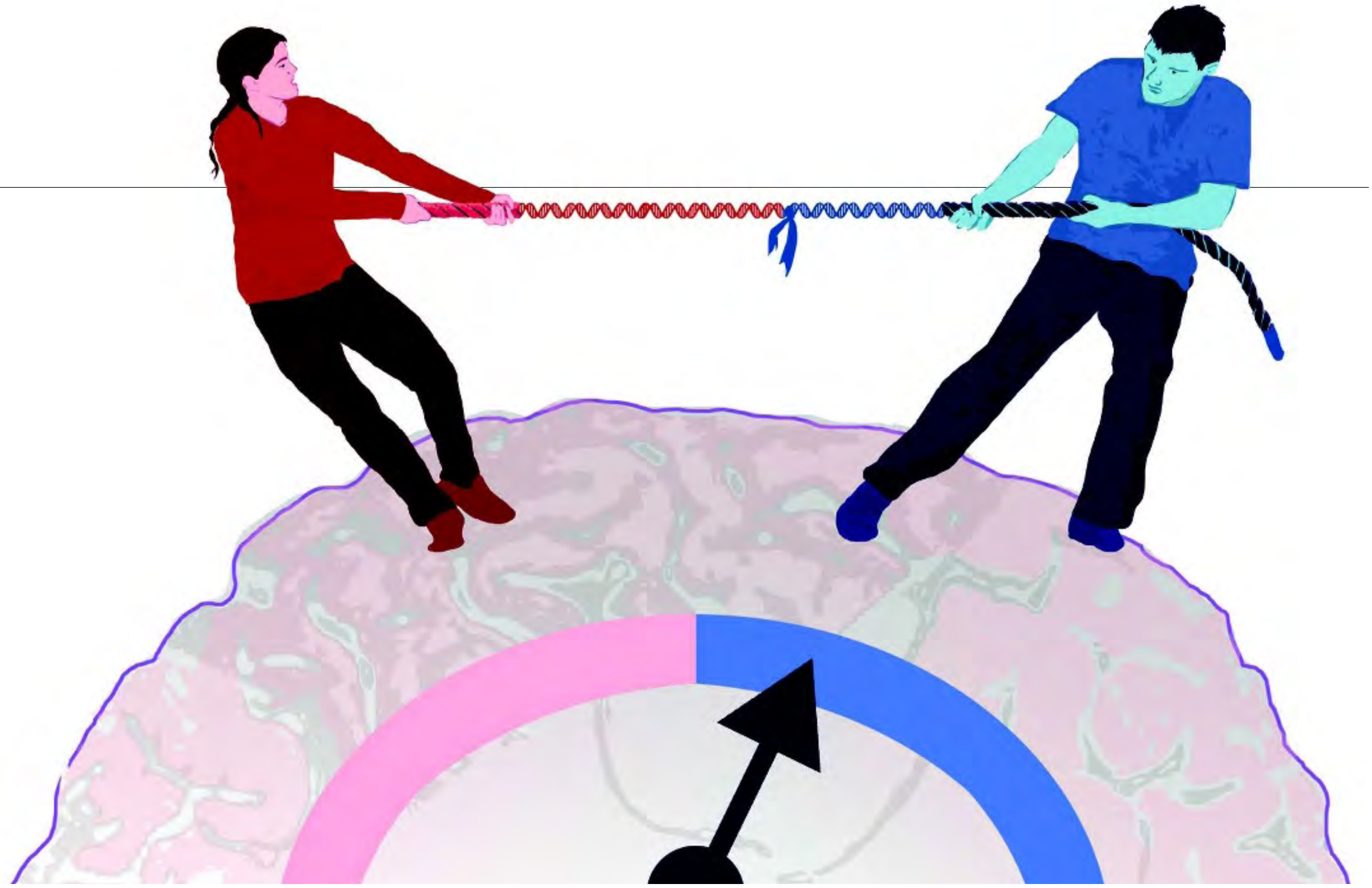
Autistic children are notably self-oriented and demanding on their care-givers; throughout evolutionary history, this has usually been the mother. This led us to suggest in 2006 that some cases of autism may be the result of paternally biased expression of genes involved with brain development[2]. Since then, investigations of patients with Beckwith–Wiedemann syndrome have revealed that they have a greatly increased risk of autism, and people with autism tend to have enhanced expression of *IGF2*. These associations support a link between autism and imprinting.

Although there is clearly a wide range of risk factors for autism, including some known toxins, genetic and hormonal factors, the vast majority of cases remain of unknown cause. Imprinting could be behind such instances.

If imprinting in one direction does indeed cause autism, then imprinting in the other direction should have some opposite effect. It struck one of us (C.B.) that some of the fundamental deficits of autism described in Simon Baron-Cohen's 1995 book *Mindblindness* contrast with classic symptoms of paranoia. For example, the defective detection of gaze seen in autism seems to be the exact opposite of paranoid delusions of being watched or spied on. The autistic inability to appreciate what goes on in groups can also be seen as the antithesis of paranoid delusions of conspiracy, which involves imagining group activity everywhere. Furthermore, the religious, magical and mystical delusions that cause people with paranoia to see evidence of mind, intention and meaning in everything seem to be the opposite of autistic deficits in theory of mind, which result in an inability to understand that others have their own beliefs and intents.

We have become convinced that autism can be considered as the diametric opposite of both paranoia and the full spectrum of related psychotic and mood disorders, which include schizophrenia, bipolar disorder and major depression. Numerous other antithetical symptoms fit this broader contrast: autistic single-mindedness versus psychotic ambivalence, autistic candour versus psychotic self-deception, autistic deficits in a sense of self versus psychotic megalomania and delusions of grandeur, and so on.

In summary, we propose that autistic spectrum conditions are characterized by deficits in theory-of-mind skills, or 'hypo-mentalism',



J. ROBINSON

whereas psychotic spectrum conditions involve the exact opposite: 'hyper-mentalism'.

According to this theory[3], small deviations in imprinted-gene expression towards a maternal bias should result in smaller babies that are energetically 'cheaper' to mothers, and who are easier behaviourally — more placid, less demanding and more mentalistically attuned to interpreting and understanding the mental states of others. Large maternally biased deviations should lead to psychosis. Conversely, small alterations towards paternal bias should lead to relatively demanding children who are more focused on 'things'; larger paternal biases should cause the severe hypo-mentalistic deficits of autism. Between these extremes would sit normal cognition.

To propose a single, overriding explanation for such a huge range of mental conditions is controversial. But there seems to be some evidence that fits, from genetics as well as epidemiology.

Evidence accrues

Geneticists have found a region of human chromosome 15, for example, that contains a set of imprinted genes. Children with a paternal bias in gene expression in this area develop Angelman or 'happy puppet' syndrome, which involves hyperactive, attention-seeking behaviour in infancy and a very high incidence of autism. Children with a maternal bias in this area develop Prader–Willi syndrome, which features extremely placid, undemanding behaviour in infancy, followed after weaning by compulsive food-seeking, which can be seen to reduce the demands on a mother. This is accompanied by rates of psychosis with depression that can approach 100%.

The theory also fits neatly with another notable contrast between autistic and psychotic spectrum conditions: age of onset. Autistic disorders typically become noticeable in childhood; psychosis mostly develops in late adolescence or early adulthood. This may be because theory-of-mind skills take many years to master, so under-development of these skills is noticed much earlier than over-development.

The gene-expression tug-of-war could also help to explain sex biases in the prevalence and severity of different mental illnesses. Recent studies by Baron-Cohen have persuasively linked autistic tendencies with exposure of the fetus to testosterone, partly explaining the considerable male bias in vulnerability to autism. But Baron-Cohen's 'extreme male brain' theory does not predict the fact that autism, although rarer in girls, is generally more severe in females. Conversely, depression is more common in females, and schizophrenia is more often severe among males.

According to our theory, there are two major axes of cognition: one determined by sex, and one by the paternal or maternal bias in gene expression. Perhaps mental disorders are more common but less severe where these two axes are compatible: in males with autism and females with depression. When the two axes are least well-matched, as in females with autism and males with psychosis, disorders seem to be much more severe.

The 'imprinted brain' hypothesis could be probed through an extensive research programme spanning studies of genes, neurodevelopment, social cognition and mental conditions. It is vital to establish a full list of the genes imprinted in the human brain, determine what these genes do, and discover how variation in their expression contributes to the development of autistic versus psychotic spectrum conditions. The search for imprinted genes is under way, and methods for finding them are improving. Still, not enough attention is focused in this direction.

Our theory of genomic conflict in brain development represents the first conceptual bridge, grounded in biology, that spans the major disorders of the social brain. If our theory proves correct, Hamilton's essential insight into genetic conflict as the fundamental driver of evolution would be vindicated, and the intricate underpinnings of mental conditions such as psychosis, or his own apparent mild autism, would be much better understood. ■

Christopher Badcock is a reader in sociology at the London School of Economics, Houghton Street, London WC2A 2AE, UK.
Bernard Crespi is professor of evolutionary biology at Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.
e-mail: crespi@sfu.ca

1. Hamilton, W. D. in *Narrow Roads of Gene Land* Vol.3: *Last Words.* 206 (W. H. Freeman/Spektrum, 2005).
2. Badcock, C. & Crespi, B. *J. Evol. Biol.* **19,** 1007–1032 (2006).
3. Crespi, B. & Badcock, C. *Behav. Brain Sci.* **31,** 241–261 (2008).

NEWS & VIEWS

HUMAN BEHAVIOUR

Share and share alike

Michael Tomasello and Felix Warneken

The happy tendency to share resources equitably — at least with members of one's own social group — is a central and unique feature of human social life. It emerges, it seems, in middle childhood.

Recent experiments[1,2] have shown that chimpanzees do not take advantage of cost-free opportunities to deliver food to other members of their group. Nor do they prevent others from getting food when they could easily do so. In most situations, our nearest primate relatives seem to be focused exclusively on the food that they themselves might get.

In the words of Fehr *et al.*[3], in a paper on page 1079 of this issue, chimpanzees seem to have no "other-regarding preferences". But human children do. In their study, Fehr *et al.* found that children aged 7 to 8 will not only provide food for others when it is of no cost to themselves, they will also seek an equal distribution, even in situations where they could potentially take a larger share for themselves (as well as in situations in which they could potentially give a larger share to the 'other'). They care about what the other is getting, and they want it to be the same as for themselves. Importantly, this is especially true when the other is from their own group (that is, is one of their classmates), rather than from a different group.

Fehr and colleagues carried out their experiments with 229 Swiss boys and girls between 3 and 8 years of age. The children were given a choice between two food-allocation options, each of which involved providing some sweets for themselves and/or some for another child (not present, but represented by a photograph). For example, in one condition they could choose between 'two for me, none for you' and 'one each'.

The authors' finding that early-school-age children value equality of resource distribution confirms previous studies[4]. What is new is the experimental design. The current study's design follows that of research in experimental economics with adults, such that the child's choice of resource distribution is anonymous (the recipient is not in the room) and it will happen only once. This is meant to rule out the possibility that, as they divvy up the rewards, the children are mainly concerned either with future reciprocity from the recipient or with their own reputation for cooperation. With these self-centred concerns ruled out, we can be more confident that the children are truly acting with the other's welfare in mind, as well as their own.

BURKE/TRIOLO PRODUCTIONS/FOODPIX

Sweet exchange: early-school-age children value equality of resource distribution.

Such other-regarding preferences are the foundation on which humans' unique forms of altruism, cooperation and social norms of fairness are built. Without a tendency to monitor and care about what others are experiencing and getting — and comparing it with what one is experiencing and getting oneself — it is difficult to imagine that human morality and culture could exist. And the tendency to prefer equal outcomes for everyone in the group is characteristic of individuals in the kinds of hunter–gatherer societies in which humans spent the vast majority of their evolutionary history, suggesting that this preference did indeed play an important part in the evolution of human cooperation[5].

The main criticism of this type of research with adults is that subjects do not really believe the assurances of anonymity — or that human psychology is such that we cannot help but operate as if we are being watched even when we 'know' we are not[6]. In Fehr and colleagues' study, the children were told that everything was anonymous and that this was a one-time event (the beneficiary 'children' were only pictured). However, the worry here is that the

NATURE|Vol 454|28 August 2008

children could easily have been concerned with what the adult experimenters — who stayed in the room as the children decided — thought of them as cooperators.

A notable result was a change to other-regarding behaviour with age: younger children (3–4 years old) were mostly self-centred, whereas the older children (7–8 years old) were concerned with equality — within their group, that is. Interestingly, this developmental pattern is different from that of another form of altruism, namely 'instrumental helping'. Human infants as young as 14–18 months readily help others do such things as fetch or stack objects, or open cabinets[7]. For this kind of helping there is direct evidence that infants are not influenced by external rewards, and indeed our nearest primate relatives sometimes help others achieve their goals as well[7,8]. So, in the case of instrumental helping, the developmental trajectory is of early altruism and other-regarding preferences, and then, presumably, growing concern with reciprocity and reputation as children learn to be selective altruists so as to avoid being exploited by others.

One difference between these studies on instrumental helping and Fehr and colleagues' research is that, in the former case, children were faced with a real individual, not pictures, and this might influence their altruism. That is, in more natural settings, where resources are being distributed among individuals who are physically present, young children might very well start out being altruistic as well. But it is also possible that altruism is not a general trait, and that there are actually two different evolutionary scenarios involved here: one for helping behaviourally and one for sharing resources. Young infants may thus be helpful behaviourally because it involves only energetic and opportunity costs. But being generous with food is another story, which might be governed especially strongly by such things as expectations of reciprocity, adult encouragement and social norms.

Fehr *et al.*[3] thus contribute to a growing collection of direct investigations of the tendency of human beings to care about and act on behalf of others. Their paper shows that, in the case of sharing resources equitably, the tendency emerges in middle childhood. It is also significant that this tendency, even in childhood, is directed mostly at others within the group, for much evidence now suggests that human cooperation is at root parochial[9]. In the end, research of this kind should help us to determine the degree to which human social life, morality and culture derive from a unique other-regarding psychology that emerges early in life. ■

Michael Tomasello and Felix Warneken are in the Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany.
e-mails: tomasello@eva.mpg.de; warneken@eva.mpg.de

1. Silk, J. B. *et al. Nature* **437,** 1357–1359 (2005).
2. Jensen, K., Hare, B., Call, J. & Tomasello, M. *Proc. R. Soc. Lond. B* **273,** 1013–1021 (2006).
3. Fehr, E., Bernhard, H. & Rockenbach, B. *Nature* **454,** 1079–1083 (2008).
4. Hook, J. G. & Cook, T. D. *Psychol. Bull.* **86,** 429–445 (1979).
5. Boehm, C. *Hierarchy in the Forest* (Harvard Univ. Press, 1999).
6. Haley, K. J. & Fessler, D. M. T. *Evol. Hum. Behav.* **26,** 245–256 (2005).
7. Warneken, F. & Tomasello, M. *Science* **311,** 1301–1303 (2006).
8. Warneken, F., Hare, B., Melis, A., Hanus, D. & Tomasello, M. *PLoS Biol.* **5,** e184 (2007).
9. Bernhard, H., Fischbacher, U. & Fehr, E. *Nature* **442,** 912–915 (2006).

EARTH SCIENCE

A sheet-metal geodynamo

Ulrich R. Christensen

A decade of modelling Earth's core on computers has led to the belief that we understand what produces Earth's magnetic field. More realistic simulations are now shaking that complacency.

Earth's magnetic field is often depicted as though a gigantic bar magnet resides inside our planet. In fact, the magnetic field is created by strong electrical currents generated by a dynamo process resulting from the flow of molten iron in Earth's outer core. An important feature of this motion is thought to be spiralling flows in the form of long columns running parallel to Earth's axis. Despite numerous simplifying assumptions, previous computer simulations of Earth's core have reproduced both this kind of flow and the properties of the magnetic field remarkably well. On page 1106 of this issue, Kageyama *et al.*[1] present a simulation that should more closely model the conditions in Earth's core. Instead of columns, the model shows thin sheets, which undermines our present thoughts on exactly how the geodynamo works.

In a simple dynamo, as might power the lights on a bicycle, the motion of a wire through a magnetic field generates electrical current in that wire. In more sophisticated, technical dynamos found in cars or power stations, the magnetic field is generated by the dynamo's own current rather than being externally applied. These dynamos work because the current is guided by wires arranged in coils. But Earth's core is an unstructured sphere of iron, which is liquid in its outer part. A simple rotational motion of the core would be insufficient to produce a magnetic field. Rather, the liquid metal must follow complex paths reminiscent of the coils in a technical dynamo.

The flow of material in the outer, liquid, core is produced by convection, driven mainly by a flux of heat and buoyant elements (such as silicon and sulphur) that emanate from the inner, solid, core. The flow is strongly influenced by forces due to Earth's rotation — such as the Coriolis force, which causes low-pressure weather systems to rotate anticlockwise in the Northern Hemisphere, and clockwise in the Southern Hemisphere. As long as convection is weak, the flow takes the form of columns, aligned north–south, through which fluid particles move with a corkscrew-like trajectory (Fig. 1). Theoretical work has shown that this situation could generate a magnetic field with a dipolar (bar-magnet-like) structure[2]. Previous computer simulations of the geodynamo that solved the equations of fluid flow and magnetic induction in a rotating sphere support the existence of such columns and their role in the generation of magnetic fields[3,4]. They show magnetic field lines emerging at the top of Earth's core mainly in so-called flux lobes, which appear in pairs aligned on the same meridian line at around 60° latitude in the Northern and Southern Hemispheres. The locations of the lobes are thought to be places where individual convection columns touch the outer boundary of the core[5].

All computer models of the geodynamo suffer from being unable to run with true core conditions, because this would require much more powerful computers than are currently available. For example, the viscosity of the liquid core must be made unrealistically large to suppress small-scale eddies in the flow that cannot be resolved on present computers. A useful measure of viscosity here is the Ekman number, which is the ratio of viscous friction to the Coriolis force. The Ekman number in the core is tiny, around 10^{-15}, but until now geodynamo simulations have only been able to operate down to Ekman numbers of 10^{-6}, with larger values being used in most cases. Kageyama *et al.*[1] used 4,096 processors at the Earth simulator in Japan to run a vigorously convecting dynamo model with an Ekman number approaching 10^{-7}.

This may seem a small step, given the discrepancy remaining — eight orders of magnitude — but such efforts are important to discover whether simulations with high Ekman numbers produce qualitatively similar results to those with more realistic conditions, or whether lowering the Ekman number results in marked alterations. It was expected that, at lower Ekman numbers, the vortex columns would become narrower and more numerous. Instead, Kageyama *et al.*[1] find that the flows become flattened into thin sheets radiating out from Earth's axis, reminiscent of the gills of a mushroom

children could easily have been concerned with what the adult experimenters — who stayed in the room as the children decided — thought of them as cooperators.

A notable result was a change to other-regarding behaviour with age: younger children (3–4 years old) were mostly self-centred, whereas the older children (7–8 years old) were concerned with equality — within their group, that is. Interestingly, this developmental pattern is different from that of another form of altruism, namely 'instrumental helping'. Human infants as young as 14–18 months readily help others do such things as fetch or stack objects, or open cabinets[7]. For this kind of helping there is direct evidence that infants are not influenced by external rewards, and indeed our nearest primate relatives sometimes help others achieve their goals as well[7,8]. So, in the case of instrumental helping, the developmental trajectory is of early altruism and other-regarding preferences, and then, presumably, growing concern with reciprocity and reputation as children learn to be selective altruists so as to avoid being exploited by others.

One difference between these studies on instrumental helping and Fehr and colleagues' research is that, in the former case, children were faced with a real individual, not pictures, and this might influence their altruism. That is, in more natural settings, where resources are being distributed among individuals who are physically present, young children might very well start out being altruistic as well. But it is also possible that altruism is not a general trait, and that there are actually two different evolutionary scenarios involved here: one for helping behaviourally and one for sharing resources. Young infants may thus be helpful behaviourally because it involves only energetic and opportunity costs. But being generous with food is another story, which might be governed especially strongly by such things as expectations of reciprocity, adult encouragement and social norms.

Fehr *et al.*[3] thus contribute to a growing collection of direct investigations of the tendency of human beings to care about and act on behalf of others. Their paper shows that, in the case of sharing resources equitably, the tendency emerges in middle childhood. It is also significant that this tendency, even in childhood, is directed mostly at others within the group, for much evidence now suggests that human cooperation is at root parochial[9]. In the end, research of this kind should help us to determine the degree to which human social life, morality and culture derive from a unique other-regarding psychology that emerges early in life. ■

Michael Tomasello and Felix Warneken are in the Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany.
e-mails: tomasello@eva.mpg.de; warneken@eva.mpg.de

1. Silk, J. B. *et al. Nature* **437,** 1357–1359 (2005).
2. Jensen, K., Hare, B., Call, J. & Tomasello, M. *Proc. R. Soc. Lond. B* **273,** 1013–1021 (2006).
3. Fehr, E., Bernhard, H. & Rockenbach, B. *Nature* **454,** 1079–1083 (2008).
4. Hook, J. G. & Cook, T. D. *Psychol. Bull.* **86,** 429–445 (1979).
5. Boehm, C. *Hierarchy in the Forest* (Harvard Univ. Press, 1999).
6. Haley, K. J. & Fessler, D. M. T. *Evol. Hum. Behav.* **26,** 245–256 (2005).
7. Warneken, F. & Tomasello, M. *Science* **311,** 1301–1303 (2006).
8. Warneken, F., Hare, B., Melis, A., Hanus, D. & Tomasello, M. *PLoS Biol.* **5,** e184 (2007).
9. Bernhard, H., Fischbacher, U. & Fehr, E. *Nature* **442,** 912–915 (2006).

EARTH SCIENCE

A sheet-metal geodynamo

Ulrich R. Christensen

A decade of modelling Earth's core on computers has led to the belief that we understand what produces Earth's magnetic field. More realistic simulations are now shaking that complacency.

Earth's magnetic field is often depicted as though a gigantic bar magnet resides inside our planet. In fact, the magnetic field is created by strong electrical currents generated by a dynamo process resulting from the flow of molten iron in Earth's outer core. An important feature of this motion is thought to be spiralling flows in the form of long columns running parallel to Earth's axis. Despite numerous simplifying assumptions, previous computer simulations of Earth's core have reproduced both this kind of flow and the properties of the magnetic field remarkably well. On page 1106 of this issue, Kageyama *et al.*[1] present a simulation that should more closely model the conditions in Earth's core. Instead of columns, the model shows thin sheets, which undermines our present thoughts on exactly how the geodynamo works.

In a simple dynamo, as might power the lights on a bicycle, the motion of a wire through a magnetic field generates electrical current in that wire. In more sophisticated, technical dynamos found in cars or power stations, the magnetic field is generated by the dynamo's own current rather than being externally applied. These dynamos work because the current is guided by wires arranged in coils. But Earth's core is an unstructured sphere of iron, which is liquid in its outer part. A simple rotational motion of the core would be insufficient to produce a magnetic field. Rather, the liquid metal must follow complex paths reminiscent of the coils in a technical dynamo.

The flow of material in the outer, liquid, core is produced by convection, driven mainly by a flux of heat and buoyant elements (such as silicon and sulphur) that emanate from the inner, solid, core. The flow is strongly influenced by forces due to Earth's rotation — such as the Coriolis force, which causes low-pressure weather systems to rotate anticlockwise in the Northern Hemisphere, and clockwise in the Southern Hemisphere. As long as convection is weak, the flow takes the form of columns, aligned north–south, through which fluid particles move with a corkscrew-like trajectory (Fig. 1). Theoretical work has shown that this situation could generate a magnetic field with a dipolar (bar-magnet-like) structure[2]. Previous computer simulations of the geodynamo that solved the equations of fluid flow and magnetic induction in a rotating sphere support the existence of such columns and their role in the generation of magnetic fields[3,4]. They show magnetic field lines emerging at the top of Earth's core mainly in so-called flux lobes, which appear in pairs aligned on the same meridian line at around 60° latitude in the Northern and Southern Hemispheres. The locations of the lobes are thought to be places where individual convection columns touch the outer boundary of the core[5].

All computer models of the geodynamo suffer from being unable to run with true core conditions, because this would require much more powerful computers than are currently available. For example, the viscosity of the liquid core must be made unrealistically large to suppress small-scale eddies in the flow that cannot be resolved on present computers. A useful measure of viscosity here is the Ekman number, which is the ratio of viscous friction to the Coriolis force. The Ekman number in the core is tiny, around 10^{-15}, but until now geodynamo simulations have only been able to operate down to Ekman numbers of 10^{-6}, with larger values being used in most cases. Kageyama *et al.*[1] used 4,096 processors at the Earth simulator in Japan to run a vigorously convecting dynamo model with an Ekman number approaching 10^{-7}.

This may seem a small step, given the discrepancy remaining — eight orders of magnitude — but such efforts are important to discover whether simulations with high Ekman numbers produce qualitatively similar results to those with more realistic conditions, or whether lowering the Ekman number results in marked alterations. It was expected that, at lower Ekman numbers, the vortex columns would become narrower and more numerous. Instead, Kageyama *et al.*[1] find that the flows become flattened into thin sheets radiating out from Earth's axis, reminiscent of the gills of a mushroom

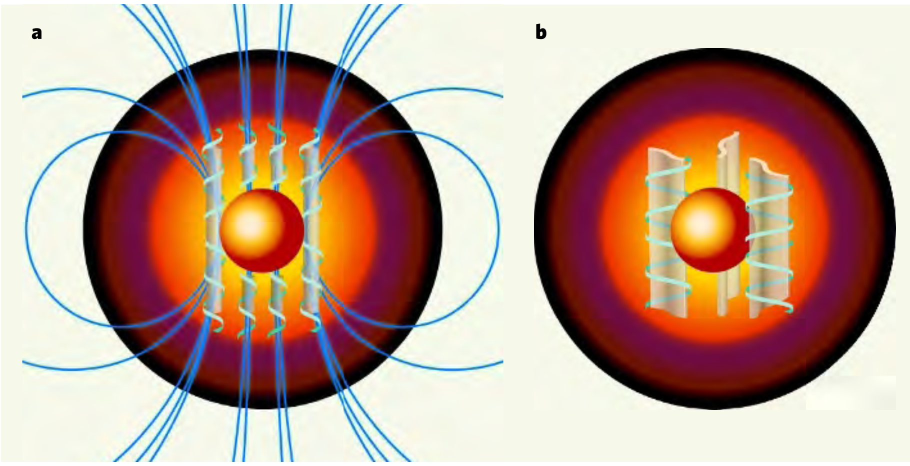


Figure 1 | Alternative geodynamos. The magnetic field is generated in Earth's liquid outer core. **a**, In the conventional concept, the molten iron circulates along a spiralling path in columns aligned in the north–south direction, generating electrical currents that set up the dipolar magnetic field. The concentration of field lines into anticyclonic vortices (rotating in the same direction as air around a region of high pressure) has been thought to explain the intense magnetic lobes found in Earth's field at the top of the core. **b**, The computer simulations of Kageyama *et al.*[1] show that the flow pattern in the core may bear more resemblance to a set of thin sheets, which puts the conventional concept into question. The sheet-like flow is efficient at generating a magnetic field, but how it might lead to the dipole geometry of Earth's field is not yet clear.

radiating away from its stem. Similar patterns have been observed in laboratory convection experiments[6]. If the flow in Earth's core has the form of elongated sheets rather than columns, the interpretation of magnetic flux lobes at the top of the core must be reconsidered.

Kageyama and colleagues' simulations[1] show that a laminated flow pattern still acts as an efficient dynamo. The conventional way of visualizing dynamo simulations is to map the resulting magnetic field lines. In complex dynamos the result often looks like a bowl of spaghetti and is hard to interpret. Kageyama *et al.* instead map lines of electrical current to produce a more coherent image, showing a large number of structures similar to the coils used in electrical engineering. However, unlike the Earth, this model produces a magnetic field that is not predominantly dipolar.

It is disturbing that dynamo models with grossly simplified parameters reproduce Earth's field well, whereas this agreement degrades when conditions are made more realistic. Some proposed explanations for this discrepancy highlight the problems encountered when performing simulations at the edge of what is currently possible. Even though the simulation took many months to run, it covered only 2,000 years of modelled time. If the flow in Earth's core were to stop suddenly, the dipole field would take 20,000 years to decay and so a similar time scale might be needed to create the field from scratch, requiring computer runs of years rather than months. In addition, the Ekman number is not the only parameter that has had to be set at unrealistic levels in geodynamo models. For example, the ratio between viscosity and electrical resistivity (magnetic Prandtl number) is set to 1 in simulations, despite being about 10^{-6} in the core.

It will remain impossible to model the geodynamo at true core conditions for the foreseeable future. The recipe for getting models with an Earth-like field might therefore be to have the parameters 'wrong in the right proportions'. The question is then what relationships between parameters must be kept fixed as we approach more realistic values. Kageyama *et al.*[1] have penetrated territory previously inaccessible to simulation. The next step will be even more demanding: scouting this new terrain to identify the true path through parameter space to understanding. ■

Ulrich R. Christensen is at the Max Planck Institute for Solar System Research, Max-Planck Strasse 2, 37191 Katlenburg-Lindau, Germany.
e-mail: christensen@mps.mpg.de

1. Kageyama, A., Miyagoshi, T. & Sato, T. *Nature* **454,** 1106–1109 (2008).
2. Busse, F. H. *Geophys. J. R. Astron. Soc.* **42,** 437–459 (1975).
3. Kageyama, A. & Sato, T. *Phys. Rev. E* **55,** 4617–4626 (1997).
4. Olson, P., Christensen, U. & Glatzmaier, G. A. *J. Geophys. Res.* **104,** 10383–10404 (1999).
5. Gubbins, D. & Bloxham, J. *Nature* **325,** 509–511 (1985).
6. Sumita, I. & Olson, P. *Phys. Earth Planet. Inter.* **117,** 153–170 (2000).

SYSTEMS BIOLOGY

Reverse engineering the cell

Nicholas T. Ingolia and Jonathan S. Weissman

Borrowing ideas that were originally developed to study electronic circuits, two reports decipher how yeast reacts to changes in its environment by analysing the organism's responses to oscillating input signals.

Systems biology has certainly caught people's attention. But when pressed to define precisely what it is, let alone how it will complement classic reductionist approaches, concrete answers can be hard to find. One immediate, practical contribution of systems biology is that it suggests new ways of interrogating cellular responses to stimuli, and provides a framework for understanding the results[1]. In this spirit, two papers[2,3] now borrow ideas from the theory of system identification — a formal strategy for building mathematical models of dynamic systems based on experimental data — to examine the response of yeast cells to changes in their environment. In both cases, the authors use microfluidics to control precisely the external stimuli and to quantify the cells' responses. The authors thus unravel some of the 'wiring' of the signal transduction networks that monitor these environmental changes.

Studies of gene regulatory networks typically measure the way in which cells respond to strong perturbations, such as changes in temperature or abrupt variations in nutrient concentration. Such marked changes are valuable for inducing signalling pathways and for identifying the molecules involved, but are not likely to be representative of physiological conditions. At worst, they might even give a distorted view of the normal function of the pathways. Fortunately, more nuanced approaches can help overcome these problems.

Engineers know that certain dynamic systems, such as electronic circuits, can be studied by measuring their response to stimuli that vary with a sinusoidal pattern[4]. In this approach, the response of the system is measured at different frequencies of periodic signals; well-developed mathematical methods are then used to convert this 'output' into a model of the inner workings of the system. The frequency dependences of two quantities are particularly informative: the amplitude of the response and the lag (phase difference) between the input stimulus and the output behaviour (Fig. 1, overleaf). This approach has previously been used to study bacterial movement in response to different frequencies of pulses of a chemical attractant[5]. But broader biological applications of the strategy have had to wait for technical advances that are only now coming into play — such as microfluidics and fluorescent cell reporters.

Enter Mettetal *et al.*[3], who report in *Science*

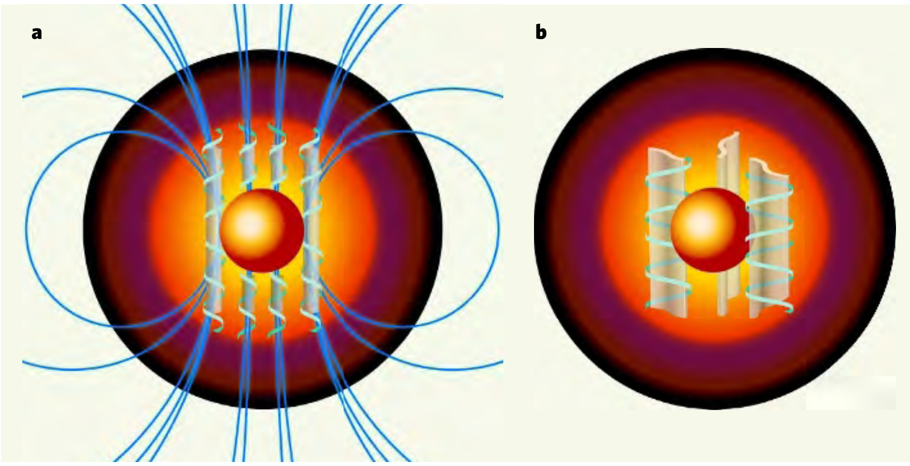


Figure 1 | Alternative geodynamos. The magnetic field is generated in Earth's liquid outer core. **a**, In the conventional concept, the molten iron circulates along a spiralling path in columns aligned in the north–south direction, generating electrical currents that set up the dipolar magnetic field. The concentration of field lines into anticyclonic vortices (rotating in the same direction as air around a region of high pressure) has been thought to explain the intense magnetic lobes found in Earth's field at the top of the core. **b**, The computer simulations of Kageyama *et al.*[1] show that the flow pattern in the core may bear more resemblance to a set of thin sheets, which puts the conventional concept into question. The sheet-like flow is efficient at generating a magnetic field, but how it might lead to the dipole geometry of Earth's field is not yet clear.

radiating away from its stem. Similar patterns have been observed in laboratory convection experiments[6]. If the flow in Earth's core has the form of elongated sheets rather than columns, the interpretation of magnetic flux lobes at the top of the core must be reconsidered.

Kageyama and colleagues' simulations[1] show that a laminated flow pattern still acts as an efficient dynamo. The conventional way of visualizing dynamo simulations is to map the resulting magnetic field lines. In complex dynamos the result often looks like a bowl of spaghetti and is hard to interpret. Kageyama *et al.* instead map lines of electrical current to produce a more coherent image, showing a large number of structures similar to the coils used in electrical engineering. However, unlike the Earth, this model produces a magnetic field that is not predominantly dipolar.

It is disturbing that dynamo models with grossly simplified parameters reproduce Earth's field well, whereas this agreement degrades when conditions are made more realistic. Some proposed explanations for this discrepancy highlight the problems encountered when performing simulations at the edge of what is currently possible. Even though the simulation took many months to run, it covered only 2,000 years of modelled time. If the flow in Earth's core were to stop suddenly, the dipole field would take 20,000 years to decay and so a similar time scale might be needed to create the field from scratch, requiring computer runs of years rather than months. In addition, the Ekman number is not the only parameter that has had to be set at unrealistic levels in geodynamo models. For example, the ratio between viscosity and electrical resistivity (magnetic Prandtl number) is set to 1 in simulations, despite being about 10^{-6} in the core.

It will remain impossible to model the geodynamo at true core conditions for the foreseeable future. The recipe for getting models with an Earth-like field might therefore be to have the parameters 'wrong in the right proportions'. The question is then what relationships between parameters must be kept fixed as we approach more realistic values. Kageyama *et al.*[1] have penetrated territory previously inaccessible to simulation. The next step will be even more demanding: scouting this new terrain to identify the true path through parameter space to understanding. ■

Ulrich R. Christensen is at the Max Planck Institute for Solar System Research, Max-Planck Strasse 2, 37191 Katlenburg-Lindau, Germany.
e-mail: christensen@mps.mpg.de

1. Kageyama, A., Miyagoshi, T. & Sato, T. *Nature* **454,** 1106–1109 (2008).
2. Busse, F. H. *Geophys. J. R. Astron. Soc.* **42,** 437–459 (1975).
3. Kageyama, A. & Sato, T. *Phys. Rev. E* **55,** 4617–4626 (1997).
4. Olson, P., Christensen, U. & Glatzmaier, G. A. *J. Geophys. Res.* **104,** 10383–10404 (1999).
5. Gubbins, D. & Bloxham, J. *Nature* **325,** 509–511 (1985).
6. Sumita, I. & Olson, P. *Phys. Earth Planet. Inter.* **117,** 153–170 (2000).

SYSTEMS BIOLOGY

Reverse engineering the cell

Nicholas T. Ingolia and Jonathan S. Weissman

Borrowing ideas that were originally developed to study electronic circuits, two reports decipher how yeast reacts to changes in its environment by analysing the organism's responses to oscillating input signals.

Systems biology has certainly caught people's attention. But when pressed to define precisely what it is, let alone how it will complement classic reductionist approaches, concrete answers can be hard to find. One immediate, practical contribution of systems biology is that it suggests new ways of interrogating cellular responses to stimuli, and provides a framework for understanding the results[1]. In this spirit, two papers[2,3] now borrow ideas from the theory of system identification — a formal strategy for building mathematical models of dynamic systems based on experimental data — to examine the response of yeast cells to changes in their environment. In both cases, the authors use microfluidics to control precisely the external stimuli and to quantify the cells' responses. The authors thus unravel some of the 'wiring' of the signal transduction networks that monitor these environmental changes.

Studies of gene regulatory networks typically measure the way in which cells respond to strong perturbations, such as changes in temperature or abrupt variations in nutrient concentration. Such marked changes are valuable for inducing signalling pathways and for identifying the molecules involved, but are not likely to be representative of physiological conditions. At worst, they might even give a distorted view of the normal function of the pathways. Fortunately, more nuanced approaches can help overcome these problems.

Engineers know that certain dynamic systems, such as electronic circuits, can be studied by measuring their response to stimuli that vary with a sinusoidal pattern[4]. In this approach, the response of the system is measured at different frequencies of periodic signals; well-developed mathematical methods are then used to convert this 'output' into a model of the inner workings of the system. The frequency dependences of two quantities are particularly informative: the amplitude of the response and the lag (phase difference) between the input stimulus and the output behaviour (Fig. 1, overleaf). This approach has previously been used to study bacterial movement in response to different frequencies of pulses of a chemical attractant[5]. But broader biological applications of the strategy have had to wait for technical advances that are only now coming into play — such as microfluidics and fluorescent cell reporters.

Enter Mettetal *et al.*[3], who report in *Science*

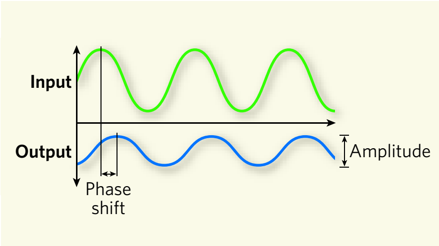


Figure 1 | Modelling dynamic systems. In systems such as electronic circuits, the input signals are transformed in a defined manner to produce the output. The output of the system in response to a sinusoidal input signal can be analysed with well-developed mathematical methods to produce a model of how the system works. The frequency dependence of two aspects of the output, namely the amplitude and the phase shift between the output and the input, is especially informative. This approach has now been used to model the biological pathways involved in the response of the yeast *S. cerevisiae* to glucose levels[2] or osmotic strength[3] in the surrounding medium.

their findings on the amplitude and phase responses of the yeast *Saccharomyces cerevisiae* to periodic changes in the osmotic strength of the surrounding medium. Using system-identification strategies borrowed from electrical engineers, they built an abstract model of the osmotic stress response from their data. They then compared this model with what was known about the molecular components of the response. This allowed them to reveal the role of different negative feedback loops in controlling the timescale of the process.

When starved of its favoured nutrient, glucose, *S. cerevisiae* can exploit an alternative sugar, galactose. In a study complementary to that of Mettetal and colleagues, Bennett *et al.*[2] (page 1119, this issue) have probed how the transcriptional regulation of the yeast's galactose-utilization genes responds to oscillations in glucose levels. These genes are induced by the presence of galactose, provided that glucose is not present.

Bennett *et al.* used microfluidics to produce a chamber that provides a controlled flow of growth medium to yeast cells. They then added glucose to the flow so that the glucose levels in the chamber varied sinusoidally. By studying cells that had been engineered to co-express a fluorescent reporter protein with one of the yeast's galactose-utilizing enzymes, the authors could quantify the cells' response to galactose. Using the microfluidics set-up, they were therefore able to measure the transcriptional response of galactose-utilizing genes to oscillations of glucose concentration at different frequencies. They found that the yeast reacts strongly to slow oscillations in glucose levels, but weakly to fast oscillations. The galactose-response pathway therefore seems to be a low-pass filter, in which cells react to long-term changes in nutrient conditions but ignore fast fluctuations.

The gene-regulatory response of yeast to galactose has been extensively studied, which allowed Bennett *et al.*[2] to use their results to make detailed, quantitative models of the biological pathway involved in the process. But their first model could not reproduce the scale of the difference between the low-frequency and high-frequency responses. The authors therefore proposed that a previously unsuspected mechanism, not included within their model, must also be involved.

Because the decay rate of messenger RNA in the pathway determines the duration of a crucial part of the model, and because glucose is known to affect the stability of some mRNAs, the authors hypothesized that regulation of mRNA stability by glucose could be the missing element of their model. They confirmed this experimentally, by showing that the mRNAs of critical galactose-response genes are destabilized by the presence of glucose. Bennett and colleagues' systems-identification strategy for studying the galactose-utilization pathway in yeast has thus led to the discovery of a previously unknown level of regulation in this well-studied process. Furthermore, the authors have linked the new regulation mechanism to a possible physiological function, namely responding to long-term changes in nutrients but not to faster fluctuations.

The authors next tested the response of a different strain of yeast — one that has a defect in its galactose-utilization pathway — to oscillations in glucose levels. This strain requires much higher levels of galactose than normal strains for steady-state activation of the relevant genes. They found that the defect affected only the phase component of the response to glucose oscillations. In other words, this strain was just as good as the normal strain at distinguishing different oscillation frequencies. Both strains are therefore capable of ignoring fast fluctuations in nutrients, but the mutant strain takes longer to respond to slow perturbations. Although otherwise accurate, Bennett and colleagues' simulations[2] of the defective yeast's response failed to predict that it would be slower to react than the normal strain. This might point to yet more unappreciated mechanistic nuances in the galactose-utilization pathway.

The experimental combination of microfluidic signal control with fluorescent output is powerful: any diffusible signal could, in principle, provide a variable stimulus, and fluorescent reporter proteins are a mainstay of modern transcriptional regulation studies. Frequency-response measurements of this sort will therefore be possible in many different biological systems. Of course, it might be that the systems-determination strategy will have limited use for interpreting the results, especially for noisy biological responses. But it should at least provide a starting point for unravelling pathways. More generally, these two papers[2,3] emphasize that the way we choose to perturb biological systems determines what aspects of those systems can be observed.

50 & 100 YEARS AGO

50 YEARS AGO

Bradenham Manor, the property of the National Trust, has been leased to the British Tabulating Machine Co., Ltd., for use as a Hollerith Computer Training Centre ... [T]he Centre was opened by Lord Halsbury ... who said that the old idea that a trainee learnt his job by copying what his predecessor could be seen to be doing has no room in the field of computer training. Training on modern, almost academic, lines is what is required ... Some 3,000 people per annum undergo training in the use and applications of mechanized accounting equipment at the five educational and training establishments maintained by the British Tabulating Machine Company's resources.

From *Nature* 30 August 1958.

100 YEARS AGO

A discussion will take place in Section D of the British Association on the abuses resulting from the strict application of the rule of priority in zoological nomenclature and on the means of protecting well-established names. Much inconvenience is caused by the extreme application of the rule ... the worst feature of which is ... the transfer of names from one to another, as we have seen in the case of Astacus, Torpedo, Holothuria, Simia, Cynocephalus ... Many zoologists think it is time to protest against the evil resulting from the indiscriminate application of what would be an excellent rule if tempered by a little consideration for tradition.

ALSO:

According to a *Times* correspondent, Dr. Lee De Forest expects that within two years Paris and New York will be in direct wireless telephonic communication. An apparatus which may ultimately transmit and receive messages to and from the Eiffel Tower is to be installed upon the 700-feet tower of the Metropolitan Life Insurance Company of New York.

From *Nature* 27 August 1908.

Technologies that permit a broader range of perturbations to be made should be welcomed, because they enable new aspects of even well-studied processes to be explored. ■

Nicholas T. Ingolia and Jonathan S. Weissman are at the Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, University of California, San Francisco, and California Institute for Quantitative Biomedical Research, San Francisco, California 94158-2542, USA.
e-mail: weissman@cmp.ucsf.edu

1. Lipan, O. & Wong, W. H. *Proc. Natl Acad. Sci. USA* **102,** 7063–7068 (2005).
2. Bennett, M. R. *et al. Nature* **454,** 1119–1122 (2008).
3. Mettetal, J. T., Muzzey D., Gómez-Uribe, C. & van Oudenaarden, A. *Science* **319,** 482–484 (2008).
4. Ogata, K. *Modern Control Engineering* (Prentice-Hall, 1970).
5. Block S. M., Segall, J. E. & Berg, H. C. *Cell* **31,** 215–226 (1982).

CONDENSED-MATTER PHYSICS

Dual realities in superconductors

Tetsuo Hanaguri

In some copper oxides, superconductivity emerges when fixed electrons become mobile. A microscopy technique reveals that this process is associated with the transfer of electrons between real and abstract spaces.

Superconductivity is a dramatic phenomenon in which a material loses all resistance to electric currents. Although the concept is simple, the electronic states that underlie superconductivity are complex, and provide a rich area for research. This is especially true for high-temperature superconductors based on copper oxide (cuprates), for which mechanisms of superconductivity remain hotly debated some two decades after their discovery.

The properties of cuprates are mainly governed by the density of mobile charge-carrying species (such as electrons). This density can be controlled by 'doping' the material with small quantities of foreign atoms. As the number of carriers is reduced, the superconductivity of the material gradually diminishes, and a strange and poorly understood electronic state — the pseudo-gap state[1] — develops. On page 1072 of this issue, Kohsaka *et al.*[2] report a startling discovery in the nature of the electronic states that underlie this process. They find that, as carrier density decreases, the delocalized electrons responsible for superconductivity are progressively replaced by localized electronic entities, which form the structure of the pseudo-gap phase.

Conventional metals have many conducting electrons, which are best thought of as being spread throughout the material in waves. For all practical purposes, these waves do not interact with one another, which means that the quantum state of the electrons can be classified using a property (known as a wavevector) that is proportional to the electron's momentum. Electrons prefer to occupy low-energy states, but Pauli's exclusion principle allows only one electron to occupy any of the quantum energy states available in the metal. As a result, the lowest energy states are filled up to a characteristic energy level, so that every electron has its own definite momentum. The boundary between occupied and unoccupied states in the abstract space of momentums defines what is called the Fermi surface.

But in some materials, the approximation that electrons do not interact with one another breaks down. This leads to many interesting phenomena, none of which can be anticipated from the properties of individual electrons. Superconductivity is one of these emergent phenomena, and involves pairs of electrons bound together by an attractive interaction. When a sufficiently large number of these 'Cooper pairs' condense into a single quantum state, superconductivity occurs. In conventional superconductors, Cooper pairs consist of two electrons travelling in opposite directions at the Fermi surface. Superconductivity is therefore a phenomenon that is best understood in the world of momentum space.

In some transition-metal compounds, including cuprates, repulsive Coulomb interactions between the negative charges on electrons also have a role. In undoped compounds, Coulomb interactions are strong enough to localize an electron at each atomic site in real space, preventing the material from conducting electricity and so forming what is known as a Mott insulator. Doping Mott insulators with atoms that have additional charge carriers often leads to exotic phenomena such as high-temperature superconductivity.

How does the electronic structure of a cuprate change in the transition from a Mott insulator to a high-temperature superconductor? To answer this question experimentally requires the application of spectroscopic techniques that probe real space and momentum space simultaneously. Various techniques have been applied to cuprates, including angle-resolved photoemission spectroscopy (ARPES, which detects electrons emitted from materials that are irradiated with light) and scanning tunnelling microscopy/spectroscopy (STM/STS, which detects electrons that hop between the surface of a material and a sharp, electrically conducting probe). But ARPES can detect only momentum space, and STM/STS can detect only real space. At least, that was the case until now.

Kohsaka *et al.*[2] have used STM/STS to probe both real space and momentum space. In the presence of imperfections in materials, electrons (as defined in momentum space) scatter and interfere. As a result, standing waves of electron density are generated in real space that can be observed with STM/STS. The wavevectors of the standing waves can be precisely determined by Fourier analysis, thus providing a measure of the electronic structure in momentum space. Because of the unconventional

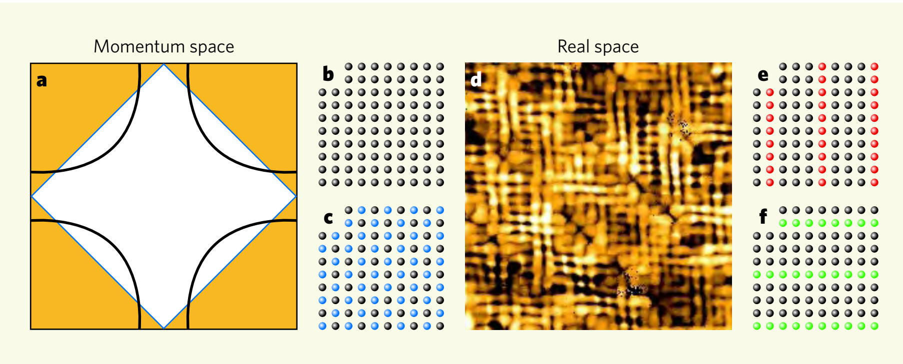


Figure 1 | Duality of electronic states in cuprates. a, The electronic states of a superconductor in momentum space are plotted here in two dimensions. The black curves denote the Fermi surface. Charges can participate in superconductivity only if they are located on regions of the Fermi surface that lie within the white zone. The yellow areas correspond to the region of momentum space in which the electrons adopt a pseudo-gap state. The blue line separating the superconducting and pseudo-gap regions is the extinction line. Kohsaka *et al.*[2] find that in a cuprate superconductor, this line does not vary with charge-carrier density. **b, c,** Assuming that the atoms of the superconductor are arranged in a grid (**b**), then a criss-cross pattern of electron states in real space (**c**, indicated by blue atoms) might explain the shape of the extinction line in momentum space. **d–f,** For the pseudo-gap state of their cuprate, Kohsaka *et al.* surprisingly observe a different pattern of electron density in real space — a mixture of horizontal and vertical domains. These might be explained by the electron periodicities shown in **e** and **f** (indicated by the green and red atoms) . But why these should be so different from the arrangement shown in **c** is a mystery.

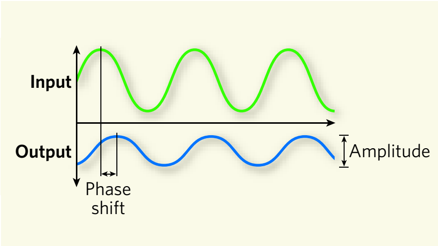


Figure 1 | Modelling dynamic systems. In systems such as electronic circuits, the input signals are transformed in a defined manner to produce the output. The output of the system in response to a sinusoidal input signal can be analysed with well-developed mathematical methods to produce a model of how the system works. The frequency dependence of two aspects of the output, namely the amplitude and the phase shift between the output and the input, is especially informative. This approach has now been used to model the biological pathways involved in the response of the yeast *S. cerevisiae* to glucose levels[2] or osmotic strength[3] in the surrounding medium.

their findings on the amplitude and phase responses of the yeast *Saccharomyces cerevisiae* to periodic changes in the osmotic strength of the surrounding medium. Using system-identification strategies borrowed from electrical engineers, they built an abstract model of the osmotic stress response from their data. They then compared this model with what was known about the molecular components of the response. This allowed them to reveal the role of different negative feedback loops in controlling the timescale of the process.

When starved of its favoured nutrient, glucose, *S. cerevisiae* can exploit an alternative sugar, galactose. In a study complementary to that of Mettetal and colleagues, Bennett *et al.*[2] (page 1119, this issue) have probed how the transcriptional regulation of the yeast's galactose-utilization genes responds to oscillations in glucose levels. These genes are induced by the presence of galactose, provided that glucose is not present.

Bennett *et al.* used microfluidics to produce a chamber that provides a controlled flow of growth medium to yeast cells. They then added glucose to the flow so that the glucose levels in the chamber varied sinusoidally. By studying cells that had been engineered to co-express a fluorescent reporter protein with one of the yeast's galactose-utilizing enzymes, the authors could quantify the cells' response to galactose. Using the microfluidics set-up, they were therefore able to measure the transcriptional response of galactose-utilizing genes to oscillations of glucose concentration at different frequencies. They found that the yeast reacts strongly to slow oscillations in glucose levels, but weakly to fast oscillations. The galactose-response pathway therefore seems to be a low-pass filter, in which cells react to long-term changes in nutrient conditions but ignore fast fluctuations.

The gene-regulatory response of yeast to galactose has been extensively studied, which allowed Bennett *et al.*[2] to use their results to make detailed, quantitative models of the biological pathway involved in the process. But their first model could not reproduce the scale of the difference between the low-frequency and high-frequency responses. The authors therefore proposed that a previously unsuspected mechanism, not included within their model, must also be involved.

Because the decay rate of messenger RNA in the pathway determines the duration of a crucial part of the model, and because glucose is known to affect the stability of some mRNAs, the authors hypothesized that regulation of mRNA stability by glucose could be the missing element of their model. They confirmed this experimentally, by showing that the mRNAs of critical galactose-response genes are destabilized by the presence of glucose. Bennett and colleagues' systems-identification strategy for studying the galactose-utilization pathway in yeast has thus led to the discovery of a previously unknown level of regulation in this well-studied process. Furthermore, the authors have linked the new regulation mechanism to a possible physiological function, namely responding to long-term changes in nutrients but not to faster fluctuations.

The authors next tested the response of a different strain of yeast — one that has a defect in its galactose-utilization pathway — to oscillations in glucose levels. This strain requires much higher levels of galactose than normal strains for steady-state activation of the relevant genes. They found that the defect affected only the phase component of the response to glucose oscillations. In other words, this strain was just as good as the normal strain at distinguishing different oscillation frequencies. Both strains are therefore capable of ignoring fast fluctuations in nutrients, but the mutant strain takes longer to respond to slow perturbations. Although otherwise accurate, Bennett and colleagues' simulations[2] of the defective yeast's response failed to predict that it would be slower to react than the normal strain. This might point to yet more unappreciated mechanistic nuances in the galactose-utilization pathway.

The experimental combination of microfluidic signal control with fluorescent output is powerful: any diffusible signal could, in principle, provide a variable stimulus, and fluorescent reporter proteins are a mainstay of modern transcriptional regulation studies. Frequency-response measurements of this sort will therefore be possible in many different biological systems. Of course, it might be that the systems-determination strategy will have limited use for interpreting the results, especially for noisy biological responses. But it should at least provide a starting point for unravelling pathways. More generally, these two papers[2,3] emphasize that the way we choose to perturb biological systems determines what aspects of those systems can be observed.

50 & 100 YEARS AGO

50 YEARS AGO

Bradenham Manor, the property of the National Trust, has been leased to the British Tabulating Machine Co., Ltd., for use as a Hollerith Computer Training Centre ... [T]he Centre was opened by Lord Halsbury ... who said that the old idea that a trainee learnt his job by copying what his predecessor could be seen to be doing has no room in the field of computer training. Training on modern, almost academic, lines is what is required ... Some 3,000 people per annum undergo training in the use and applications of mechanized accounting equipment at the five educational and training establishments maintained by the British Tabulating Machine Company's resources.

From *Nature* 30 August 1958.

100 YEARS AGO

A discussion will take place in Section D of the British Association on the abuses resulting from the strict application of the rule of priority in zoological nomenclature and on the means of protecting well-established names. Much inconvenience is caused by the extreme application of the rule ... the worst feature of which is ... the transfer of names from one to another, as we have seen in the case of Astacus, Torpedo, Holothuria, Simia, Cynocephalus ... Many zoologists think it is time to protest against the evil resulting from the indiscriminate application of what would be an excellent rule if tempered by a little consideration for tradition.

ALSO:

According to a *Times* correspondent, Dr. Lee De Forest expects that within two years Paris and New York will be in direct wireless telephonic communication. An apparatus which may ultimately transmit and receive messages to and from the Eiffel Tower is to be installed upon the 700-feet tower of the Metropolitan Life Insurance Company of New York.

From *Nature* 27 August 1908.

Technologies that permit a broader range of perturbations to be made should be welcomed, because they enable new aspects of even well-studied processes to be explored. ■

Nicholas T. Ingolia and Jonathan S. Weissman are at the Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, University of California, San Francisco, and California Institute for Quantitative Biomedical Research, San Francisco, California 94158-2542, USA.
e-mail: weissman@cmp.ucsf.edu

1. Lipan, O. & Wong, W. H. *Proc. Natl Acad. Sci. USA* **102,** 7063–7068 (2005).
2. Bennett, M. R. *et al. Nature* **454,** 1119–1122 (2008).
3. Mettetal, J. T., Muzzey D., Gómez-Uribe, C. & van Oudenaarden, A. *Science* **319,** 482–484 (2008).
4. Ogata, K. *Modern Control Engineering* (Prentice-Hall, 1970).
5. Block S. M., Segall, J. E. & Berg, H. C. *Cell* **31,** 215–226 (1982).

CONDENSED-MATTER PHYSICS

Dual realities in superconductors

Tetsuo Hanaguri

In some copper oxides, superconductivity emerges when fixed electrons become mobile. A microscopy technique reveals that this process is associated with the transfer of electrons between real and abstract spaces.

Superconductivity is a dramatic phenomenon in which a material loses all resistance to electric currents. Although the concept is simple, the electronic states that underlie superconductivity are complex, and provide a rich area for research. This is especially true for high-temperature superconductors based on copper oxide (cuprates), for which mechanisms of superconductivity remain hotly debated some two decades after their discovery.

The properties of cuprates are mainly governed by the density of mobile charge-carrying species (such as electrons). This density can be controlled by 'doping' the material with small quantities of foreign atoms. As the number of carriers is reduced, the superconductivity of the material gradually diminishes, and a strange and poorly understood electronic state — the pseudo-gap state[1] — develops. On page 1072 of this issue, Kohsaka *et al.*[2] report a startling discovery in the nature of the electronic states that underlie this process. They find that, as carrier density decreases, the delocalized electrons responsible for superconductivity are progressively replaced by localized electronic entities, which form the structure of the pseudo-gap phase.

Conventional metals have many conducting electrons, which are best thought of as being spread throughout the material in waves. For all practical purposes, these waves do not interact with one another, which means that the quantum state of the electrons can be classified using a property (known as a wavevector) that is proportional to the electron's momentum. Electrons prefer to occupy low-energy states, but Pauli's exclusion principle allows only one electron to occupy any of the quantum energy states available in the metal. As a result, the lowest energy states are filled up to a characteristic energy level, so that every electron has its own definite momentum. The boundary between occupied and unoccupied states in the abstract space of momentums defines what is called the Fermi surface.

But in some materials, the approximation that electrons do not interact with one another breaks down. This leads to many interesting phenomena, none of which can be anticipated from the properties of individual electrons. Superconductivity is one of these emergent phenomena, and involves pairs of electrons bound together by an attractive interaction. When a sufficiently large number of these 'Cooper pairs' condense into a single quantum state, superconductivity occurs. In conventional superconductors, Cooper pairs consist of two electrons travelling in opposite directions at the Fermi surface. Superconductivity is therefore a phenomenon that is best understood in the world of momentum space.

In some transition-metal compounds, including cuprates, repulsive Coulomb interactions between the negative charges on electrons also have a role. In undoped compounds, Coulomb interactions are strong enough to localize an electron at each atomic site in real space, preventing the material from conducting electricity and so forming what is known as a Mott insulator. Doping Mott insulators with atoms that have additional charge carriers often leads to exotic phenomena such as high-temperature superconductivity.

How does the electronic structure of a cuprate change in the transition from a Mott insulator to a high-temperature superconductor? To answer this question experimentally requires the application of spectroscopic techniques that probe real space and momentum space simultaneously. Various techniques have been applied to cuprates, including angle-resolved photoemission spectroscopy (ARPES, which detects electrons emitted from materials that are irradiated with light) and scanning tunnelling microscopy/spectroscopy (STM/STS, which detects electrons that hop between the surface of a material and a sharp, electrically conducting probe). But ARPES can detect only momentum space, and STM/STS can detect only real space. At least, that was the case until now.

Kohsaka *et al.*[2] have used STM/STS to probe both real space and momentum space. In the presence of imperfections in materials, electrons (as defined in momentum space) scatter and interfere. As a result, standing waves of electron density are generated in real space that can be observed with STM/STS. The wavevectors of the standing waves can be precisely determined by Fourier analysis, thus providing a measure of the electronic structure in momentum space. Because of the unconventional

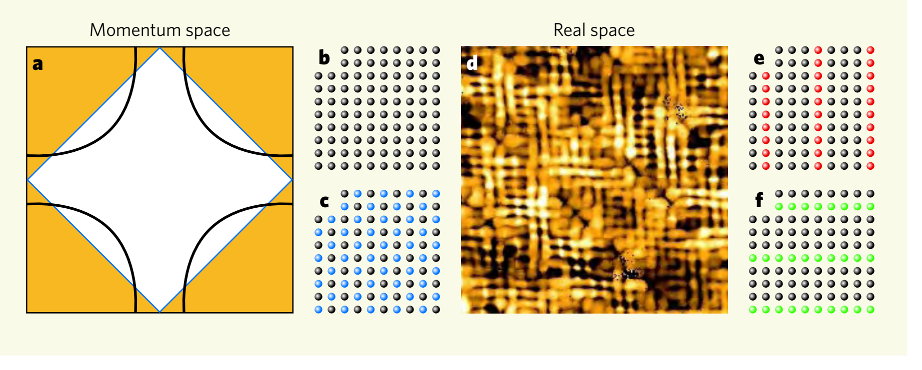


Figure 1 | Duality of electronic states in cuprates. a, The electronic states of a superconductor in momentum space are plotted here in two dimensions. The black curves denote the Fermi surface. Charges can participate in superconductivity only if they are located on regions of the Fermi surface that lie within the white zone. The yellow areas correspond to the region of momentum space in which the electrons adopt a pseudo-gap state. The blue line separating the superconducting and pseudo-gap regions is the extinction line. Kohsaka *et al.*[2] find that in a cuprate superconductor, this line does not vary with charge-carrier density. **b**, **c**, Assuming that the atoms of the superconductor are arranged in a grid (**b**), then a criss-cross pattern of electron states in real space (**c**, indicated by blue atoms) might explain the shape of the extinction line in momentum space. **d**–**f**, For the pseudo-gap state of their cuprate, Kohsaka *et al.* surprisingly observe a different pattern of electron density in real space — a mixture of horizontal and vertical domains. These might be explained by the electron periodicities shown in **e** and **f** (indicated by the green and red atoms). But why these should be so different from the arrangement shown in **c** is a mystery.

DEVELOPMENTAL GENETICS

A sex-specific switch


BLICKWINKEL/ALAMY

Those who are intimate with the fruitfly *Drosophila melanogaster* will recognize this fine specimen as a male. Most notably, the last two of its abdominal segments are pigmented. The equivalent segments in the posterior abdomen of females lack pigment, and sexual dimorphism of this and many other forms is responsible for all sorts of male–female behavioural differences in all sorts of organisms — vertebrate and invertebrate. Thomas Williams and colleagues have dissected this instance of fruitfly sexual dimorphism in molecular detail to reveal the genetic switch concerned (T. M. Williams *et al. Cell* **134,** 610–623; 2008).

The pigmentation is regulated by the tandem duplicate *bab* genes, whose protein products repress the enzyme system that produces pigment. The authors' investigations centred on their discovery of two *cis*-regulatory elements that respectively control *bab* expression in the anterior and posterior segments, these elements being themselves binding sites for transcription factors. One of the transcription factors is ABD-B, a member of the famed HOX family; two others are forms of the Doublesex protein, one specific for males, the other for females.

The sex-specific differences in the posterior segments arise from different control of the *cis*-regulatory element. In females, a combination of the transcription factors ABD-B and the female-specific form of Doublesex activates *bab* expression (and so prevents pigmentation). In males, the male-specific form of Doublesex overrides the activating effect of ABD-B: *bab* is repressed, so unleashing the enzyme system that pigments the final two male segments.

The most intricate part of the authors' work, however, comes in their studies of the evolution of this switch. From comparisons with transcription-factor binding sites on the equivalent *cis*-regulatory element of another species of *Drosophila*, *D. willistoni*, they aimed to identify the molecular changes that occurred in the lineage of *D. melanogaster* arising from a putative monomorphic common ancestor. The authors' conclusion is that rather than a wholesale gain of ABD-B or Doublesex binding sites, the advent of male pigmentation was due to a suite of fine-scale changes in the number, polarity and topology of sites.

Williams *et al.* don't go into the possible evolutionary and ecological factors that drove the production of male pigmentation. Instead, they end with the thought that, in evolutionary history, some of the main differences in the body plans of arthropods, like *Drosophila,* and vertebrates have arisen from alterations in the pattern of gene expression along the principal body axis. Their work, they say, shows a general way in which such expression patterns could evolve through the accumulation of many fine-scale alterations to a gene's *cis*-regulatory elements.

Tim Lincoln

nature of electron pairing in cuprates, the binding energy of the Cooper pair is highly anisotropic, and vanishes in certain directions (known as nodal directions) in momentum space. The wave-properties of electrons excited from such a nodal, superconducting state produce unique scattering interference patterns[3–5]. Electronic standing waves in real space are therefore a good probe of superconductivity in momentum space.

Kohsaka *et al.*[2] performed Fourier-transform STM/STS on crystals of the cuprate superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$. They observed standing waves associated with the nodal superconductivity for low-energy electrons, but these diminished for electrons that had energy greater than a certain 'extinction' value. By analysing these data, the authors identified the segment of the Fermi surface that supports superconductivity. Their measurements on a selection of cuprate crystals that have different carrier densities showed that the extinction energy is only weakly dependent on the density of carriers, but that the superconducting segment of the Fermi surface shrinks with decreasing carrier density.

Each superconducting segment terminates near a specific line in momentum space: the extinction line. Boundaries in momentum space are often associated with periodicities in real space. For example, the periodicity of a crystal's atomic lattice produces boundaries in momentum space that partition that space into small regions, known as Brillouin zones (Fig. 1). The extinction line of $Bi_2Sr_2CaCu_2O_{8+\delta}$ forms a diamond shape that connects the midpoints of the Brillouin zone boundaries.

Using the real-world imaging capabilities of STM/STS, Kohsaka *et al.* observed 'glassy' electronic domains at higher electron energies, beyond the extinction line. These domains locally break the symmetry of the underlying atomic lattice (Fig. 1d) and had previously been discovered by the same group in similar, lower-resolution measurements[6]. But in their latest experiments[2], the researchers were able to sufficiently resolve the energy of the electrons to show that these glassy domains are most prominent at the characteristic energy of the pseudo-gap state. The electronic state of cuprates is thus duplicitous: it adopts either a low-energy superconducting state supported by a segment of the Fermi surface, or a high-energy pseudo-gap state, characterized by unusual real-space structures.

The observations of Kohsaka *et al.* provide definitive proof of the intricate relationship between real and momentum spaces in cuprates, and suggest several exciting directions for future research. For starters, the cause of the diagonal extinction line needs to be studied in more detail. As the real-space domains observed at the pseudo-gap energy break the symmetry of the underlying atomic lattices and have a characteristic width[6], the appearance of an associated new boundary in momentum space is not strange in itself. But the direction of the boundary associated with the observed real-space structure in the pseudo-gap state differs markedly from that of the measured extinction line, which defines the superconducting segment in momentum space (Fig. 1). This might have profound implications for the nature of the pseudo-gap and, by extension, for the mechanism of superconductivity in cuprates.

The effects of a wider range of dopings also remain to be explored — for example, those that bridge the pseudo-gap and Mott insulator states. And it will be necessary to repeat direct momentum-space measurements using ARPES. These should not only reinforce the current data, but will also make clear the differences and similarities between ARPES and the emerging technique of Fourier-transform STM/STS. ■

Tetsuo Hanaguri is in the Magnetic Materials Laboratory, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.
e-mail: hanaguri@riken.jp

1. Timusk, T. & Statt, B. *Rep. Prog. Phys.* **62,** 61–122 (1999).
2. Kohsaka, Y. *et al. Nature* **454,** 1072–1078 (2008).
3. Hoffman, J. E. *et al. Science* **297,** 1148–1151 (2002).
4. McElroy, K. *et al. Nature* **422,** 592–596 (2003).
5. Hanaguri, T. *et al. Nature Phys.* **3,** 865–871 (2007).
6. Kohsaka, Y. *et al. Science* **315,** 1380–1385 (2007).

DEVELOPMENTAL GENETICS

A sex-specific switch

Those who are intimate with the fruitfly *Drosophila melanogaster* will recognize this fine specimen as a male. Most notably, the last two of its abdominal segments are pigmented. The equivalent segments in the posterior abdomen of females lack pigment, and sexual dimorphism of this and many other forms is responsible for all sorts of male–female behavioural differences in all sorts of organisms — vertebrate and invertebrate. Thomas Williams and colleagues have dissected this instance of fruitfly sexual dimorphism in molecular detail to reveal the genetic switch concerned (T. M. Williams *et al. Cell* **134,** 610–623; 2008).

The pigmentation is regulated by the tandem duplicate *bab* genes, whose protein products repress the enzyme system that produces pigment. The authors' investigations centred on their discovery of two *cis*-regulatory elements that respectively control *bab* expression in the anterior and posterior segments, these elements being themselves binding sites for transcription factors. One of the transcription factors is ABD-B, a member of the famed HOX family; two others are forms of the Doublesex protein, one specific for males, the other for females.

The sex-specific differences in the posterior segments arise from different control of the *cis*-regulatory element. In females, a combination of the transcription factors ABD-B and the female-specific form of Doublesex activates *bab* expression (and so prevents pigmentation). In males, the male-specific form of Doublesex overrides the activating effect of ABD-B: *bab* is repressed, so unleashing the enzyme system that pigments the final two male segments.

The most intricate part of the authors' work, however, comes in their studies of the evolution of this switch. From comparisons with transcription-factor binding sites on the equivalent *cis*-regulatory element of another species of *Drosophila*, *D. willistoni*, they aimed to identify the molecular changes that occurred in the lineage of *D. melanogaster* arising from a putative monomorphic common ancestor. The authors' conclusion is that rather than a wholesale gain of ABD-B or Doublesex binding sites, the advent of male pigmentation was due to a suite of fine-scale changes in the number, polarity and topology of sites.


BLICKWINKEL/ALAMY

Williams *et al.* don't go into the possible evolutionary and ecological factors that drove the production of male pigmentation. Instead, they end with the thought that, in evolutionary history, some of the main differences in the body plans of arthropods, like *Drosophila,* and vertebrates have arisen from alterations in the pattern of gene expression along the principal body axis. Their work, they say, shows a general way in which such expression patterns could evolve through the accumulation of many fine-scale alterations to a gene's *cis*-regulatory elements.

Tim Lincoln

nature of electron pairing in cuprates, the binding energy of the Cooper pair is highly anisotropic, and vanishes in certain directions (known as nodal directions) in momentum space. The wave-properties of electrons excited from such a nodal, superconducting state produce unique scattering interference patterns[3–5]. Electronic standing waves in real space are therefore a good probe of superconductivity in momentum space.

Kohsaka *et al.*[2] performed Fourier-transform STM/STS on crystals of the cuprate superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$. They observed standing waves associated with the nodal superconductivity for low-energy electrons, but these diminished for electrons that had energy greater than a certain 'extinction' value. By analysing these data, the authors identified the segment of the Fermi surface that supports superconductivity. Their measurements on a selection of cuprate crystals that have different carrier densities showed that the extinction energy is only weakly dependent on the density of carriers, but that the superconducting segment of the Fermi surface shrinks with decreasing carrier density.

Each superconducting segment terminates near a specific line in momentum space: the extinction line. Boundaries in momentum space are often associated with periodicities in real space. For example, the periodicity of a crystal's atomic lattice produces boundaries in momentum space that partition that space into small regions, known as Brillouin zones (Fig. 1). The extinction line of $Bi_2Sr_2CaCu_2O_{8+\delta}$ forms a diamond shape that connects the midpoints of the Brillouin zone boundaries.

Using the real-world imaging capabilities of STM/STS, Kohsaka *et al.* observed 'glassy' electronic domains at higher electron energies, beyond the extinction line. These domains locally break the symmetry of the underlying atomic lattice (Fig. 1d) and had previously been discovered by the same group in similar, lower-resolution measurements[6]. But in their latest experiments[2], the researchers were able to sufficiently resolve the energy of the electrons to show that these glassy domains are most prominent at the characteristic energy of the pseudo-gap state. The electronic state of cuprates is thus duplicitous: it adopts either a low-energy superconducting state supported by a segment of the Fermi surface, or a high-energy pseudo-gap state, characterized by unusual real-space structures.

The observations of Kohsaka *et al.* provide definitive proof of the intricate relationship between real and momentum spaces in cuprates, and suggest several exciting directions for future research. For starters, the cause of the diagonal extinction line needs to be studied in more detail. As the real-space domains observed at the pseudo-gap energy break the symmetry of the underlying atomic lattices and have a characteristic width[6], the appearance of an associated new boundary in momentum space is not strange in itself. But the direction of the boundary associated with the observed real-space structure in the pseudo-gap state differs markedly from that of the measured extinction line, which defines the superconducting segment in momentum space (Fig. 1). This might have profound implications for the nature of the pseudo-gap and, by extension, for the mechanism of superconductivity in cuprates.

The effects of a wider range of dopings also remain to be explored — for example, those that bridge the pseudo-gap and Mott insulator states. And it will be necessary to repeat direct momentum-space measurements using ARPES. These should not only reinforce the current data, but will also make clear the differences and similarities between ARPES and the emerging technique of Fourier-transform STM/STS. ■

Tetsuo Hanaguri is in the Magnetic Materials Laboratory, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.
e-mail: hanaguri@riken.jp

1. Timusk, T. & Statt, B. *Rep. Prog. Phys.* **62,** 61–122 (1999).
2. Kohsaka, Y. *et al. Nature* **454,** 1072–1078 (2008).
3. Hoffman, J. E. *et al. Science* **297,** 1148–1151 (2002).
4. McElroy, K. *et al. Nature* **422,** 592–596 (2003).
5. Hanaguri, T. *et al. Nature Phys.* **3,** 865–871 (2007).
6. Kohsaka, Y. *et al. Science* **315,** 1380–1385 (2007).

REVIEWS

Puzzles, promises and a cure for ageing

Jan Vijg[1]† & Judith Campisi[1,2]

Recent discoveries in the science of ageing indicate that lifespan in model organisms such as yeast, nematodes, flies and mice is plastic and can be manipulated by genetic, nutritional or pharmacological intervention. A better understanding of the targets of such interventions, as well as the proximate causes of ageing-related degeneration and disease, is essential before we can evaluate if abrogation of human senescence is a realistic prospect.

The inevitability of ageing and death has preoccupied humanity for more than 5,000 years. In the epic named after him, Gilgamesh, the Sumerian king of Uruk, seeks to escape death, but ultimately concludes this is futile and turns to lasting works of culture to achieve immortality. Nowadays, disease—not death and the opposite, eternal life—preoccupies most biomedical scientists. Nevertheless, advances in understanding basic ageing mechanisms have made it difficult to ignore the issue of whether biomedical interventions to postpone ageing substantially are scientifically plausible. The overarching question is not whether mean human lifespan will increase modestly over the next decades. It almost certainly will, assuming continued success in reducing old-age morbidity and mortality[1]. Rather, the issue is whether postponing human ageing and natural death for many decades, possibly indefinitely, is feasible. Here, we discuss what we do and do not know about the mechanisms of ageing, and whether we have sufficient knowledge to foresee substantially retarding, halting or even reversing the degeneration and disease that limit human lifespan.

Lifespan is plastic

Ageing research has suffered more than disease-oriented research from unsubstantiated claims of potential cures. The field is also rife with opposing claims—that human life cannot be extended beyond a soft limit (120–125 years). For example, predictions in 1990 claimed that declines in death rates would not reach levels required for life expectancy at birth to exceed 85 years (ref. 2). However, Japanese females have already surpassed this limit (http://miranda.sourceoecd.org/vl=2231472/cl=20/nw=1/rpsv/health2007/g2-1-02.htm), and life expectancy in developed countries is now predicted to exceed 85 years by the year 2050 (ref. 1). A question remains as to whether science can free us from the bonds that seem to fix our lifespan.

With the discovery in the 1980s that mutations in single genes can significantly extend lifespan in the nematode *Caenorhabditis elegans*[3,4], ageing began to be viewed as malleable by methods used to understand and manipulate development and disease. At present, hundreds of mutant genes can increase longevity in model organisms, including nematodes, yeast (*Saccharomyces cerevisiae*), fruitflies (*Drosophila melanogaster*) and mice (*Mus musculus*). Most act in evolutionarily conserved pathways that regulate growth, energy metabolism, nutrition sensing and/or reproduction[5]. Examples include genes encoding components of the insulin/insulin-like growth factor 1 (IGF-I) signalling (IIS) pathway, the target of rapamycin (TOR) pathway, and the mitochondrial electron transport chain (Fig. 1). In most cases, lifespan extension occurs when activity of the component is diminished. This abatement is thought to reduce somatic damage and/or increase somatic maintenance and repair[5]. Most pro-longevity mutations are discovered by mutagen or RNA interference screens, which mainly uncover inactivated or diminished gene functions, so further longevity mutations, resulting from enhanced gene function, may yet be discovered.

Many pro-longevity mutations mimic dietary restriction—underfeeding without malnutrition—which was shown to extend lifespan in laboratory rodents almost a century ago[6]. Dietary restriction increases longevity in many species, including yeast, nematodes, spiders and dogs. Although dietary restriction phenotypes often overlap with those conferred by dampening pro-ageing pathways, in some cases dietary restriction synergizes with pro-longevity mutations, indicating that dietary restriction can act independently[7].

The appreciation that lifespan is plastic and under negative influence by genes that favour growth or procreation fuels hopes of finding small molecules that target the pathways affected by dietary restriction or pro-longevity mutations[8,9]. Compounds have been identified that show promise in this regard (Box 1), although none has yet shown major effects on lifespan in healthy mice[10]. Most, however, are in clinical trials to treat ageing-related diseases, such as diabetes and cancer, and already had the attention of the pharmaceutical industry. Of note, the US Food and Drug Administration will not approve agents that merely impede ageing. Nevertheless, it is remarkable that barely two decades ago human lifespan extension was a fantasy, whereas now pharmacological strategies are being considered for exactly that purpose. For some, the discovery that mutations in evolutionarily conserved pathways can greatly extend lifespan in model organisms is a step towards curing ageing in humans. As we argue, this viewpoint may be premature.

Diminishing returns of complexity and idiosyncratic models

There is ample evidence that much of basic biology is similar across species as divergent as yeast and humans[11]. This is also true for ageing. Studies of yeast, nematodes and flies undoubtedly illuminate our understanding of the evolutionary and mechanistic bases of human ageing. Still, the response of simple organisms to interventions might not be predictive when complexity increases, or when physiology deviates significantly from humans.

The impact of complexity is illustrated by the IIS pathway. Invertebrates have a single receptor that binds ligands similar to insulin or IGF-1. Mutations that partially blunt signalling from this receptor extend lifespan in nematodes and flies[5]. However, mammals have distinct receptors for insulin and IGF-1, with different but overlapping functions. IGF-I primarily controls growth, whereas insulin regulates metabolism. In mammals, defective insulin signalling causes insulin

[1]Buck Institute for Age Research, 8001 Redwood Boulevard, Novato, California 94945, USA. [2]Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mailstop 84–171, Berkeley, California 94720, USA. †Present address: Department of Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA.

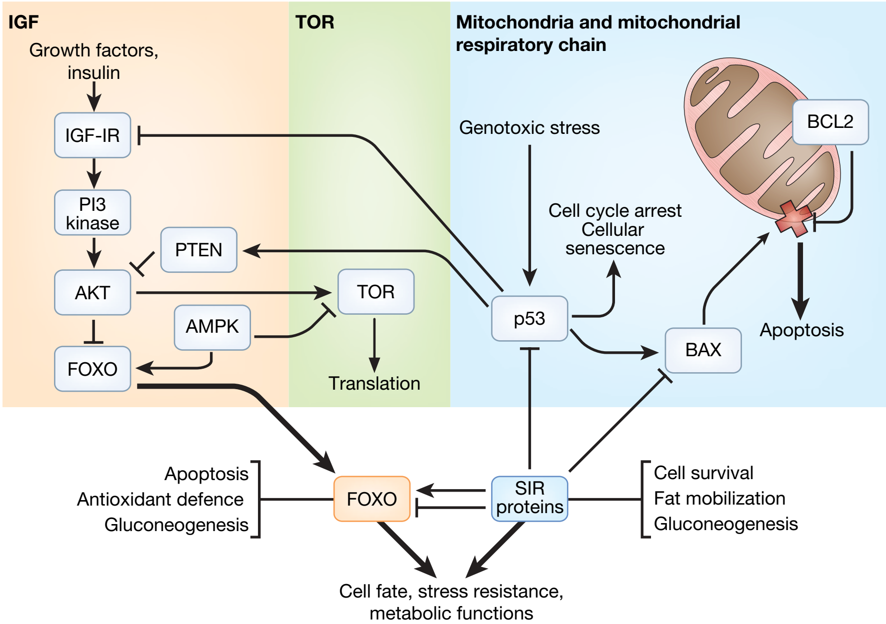


Figure 1 | Potentially conserved pro-ageing pathways, their interconnections and possible targets for intervention. In this very simplified depiction, three main pathways, the IIS, TOR and mitochondrial pathway, are indicated. The pro-ageing activities of these pathways are conserved across species, with energy sensors, such as AMPK, as potentially important hubs in the complex networks that integrate them. However, it is important to note potential dissimilarities among species as well. Most, if not all, defects in the mitochondrial respiratory chain are lethal or cause disease in humans[62], but can increase lifespan in nematodes or yeast. In mammals, mitochondria play an important part in signalling apoptosis, which can either drive or retard ageing, depending on the cell type. There is evidence that many longevity signals converge on members of the FOXO and sirtuin protein families, which can interact. Note that SIR proteins can both activate and repress FOXO. Moreover, the effects of FOXO and SIR2 in cells can be either beneficial (for example, increasing antioxidant defence) or detrimental (for example, apoptosis), and may or may not promote organismal survival. For example, in mammals, SIRT1 dampens apoptosis by repressing FOXO, but also by repressing BAX activation, thereby preventing its oligomerization into the mitochondria outer membrane (cross), which normally triggers permeabilization of the membrane and release of soluble apoptogenic factors, such as cytochrome *c*, into the cytosol. Apoptosis can be beneficial, for example, by eliminating damaged cells and preventing cancer, or can be detrimental, by eliminating irreplaceable cells, such as neurons.

resistance and diabetes. Defective IGF-I signalling causes protein breakdown and muscle degeneration. Indeed, IGF-I overexpression reduces ageing-associated cardiac dysfunction[12] and improves muscle regeneration[13]. Nonetheless, reduced insulin signalling specifically in adipose tissue, or reduced IGF-I signalling throughout the animal, modestly extends lifespan in mice[14,15]. So, tissue-specific modulation of particular signalling pathways might retard ageing in humans.

The impact of complexity is also illustrated by interactions of two longevity-modulating protein families: forkhead (FOXO) transcription factors and silent information regulator (SIR) protein deacetylases (sirtuins). FOXO proteins (DAF-16 in nematodes) are required for the lifespan extension conferred by IIS mutations, and overexpression of SIR2 orthologues increases lifespan in yeast, nematodes and flies[5]. Mammals have multiple FOXO and SIR proteins. Some FOXO proteins initiate stress-induced cell death (apoptosis), which eliminates damaged or dysfunctional cells. FOXO proteins also upregulate antioxidant defence and DNA repair-facilitating genes[16]. FOXO is deacetylated by the human SIR2 orthologue SIRT1, which increases FOXO-dependent resistance to oxidative stress and cell-cycle arrest, but inhibits FOXO-dependent apoptosis[17]. SIRT1 also deacetylates the p53 tumour suppressor protein, attenuating its transcriptional activity and suppressing stress-induced apoptosis and cellular senescence (irreversible cell cycle arrest)[18].

A priori, it is difficult to predict whether increased FOXO or sirtuin activity should increase or decrease mammalian longevity. Mammals rely on apoptosis and senescence to suppress cancer[19], which is a major age-related disease and challenge to the longevity of mammals. This is not the case for yeast, flies and nematodes, which do not or rarely develop cancer because they are unicellular or as adults they contain largely post-mitotic somatic cells.

The magnitude of longevity extensions is worth noting. Genetic dampening of IIS routinely increases lifespan in nematodes twofold[5], and one mutation increases lifespan tenfold[20]. In flies, however, single-gene mutations in the IIS or other pro-ageing pathways generally extend lifespan by only 25%–30%[21]. In mice, inactivating mutations in the *Pou1f1*, *Prop1*, or growth hormone receptor genes (which reduce IGF-I signals) increase lifespan by 40%[22], whereas mutations that directly affect IIS extend lifespan by 20% or less[14,15].

Thus, reduced IIS can substantially increase lifespan in nematodes, but much less so in the more complex fly and mouse. We know very little about mechanisms responsible for these species-specific differences. Within a species, genetic background, environment and sex differences matter. For example, lifespan extension in the IIS mutant fly *chico* depends on food concentration[23]. In transgenic flies overexpressing human superoxide dismutase in motor neurons, longevity benefits varied considerably among ten wild-derived genotypes, as well as by sex[24].

Although it is possible that we have not yet defined optimal conditions for downregulating conserved ageing pathways in organisms more complex than nematodes, other characteristics set nematodes apart. First, many of the initial longevity mutations identified in nematodes affect an alternative developmental stage termed dauer, which suspends reproduction and alters metabolism. Hibernation, which temporarily suspends mammalian metabolism and reproduction, never lasts longer than the lifespan, in contrast to the dauer state in nematodes. Second, aerobic respiration is less critical for nematodes than for flies and mammals. This may explain why RNA interference screens for increased lifespan in nematodes identified multiple genes encoding mitochondrial respiratory chain subunits[25]. Downregulating these genes may increase lifespan by

Box 1 | Strategies to counter intrinsic ageing

Evidence is emerging that identified pro-ageing pathways can be manipulated to extend lifespan by pharmacological means, although the mechanism by which such interventions act is not always clear. Small polyphenolics such as resveratrol and fisiten increase SIR2 deacetylase activity *in vitro*, and increase lifespan in a Sir2-dependent manner in yeast, *C. elegans* and *Drosophila*[9]. However, resveratrol may also be an antioxidant and suppress oxidative damage. Resveratrol increases insulin sensitivity and the survival of mice fed a high caloric diet[64]. The antifungal antibiotic rapamycin inhibits the pro-ageing TOR pathway in yeast and human cells. It is not known if TOR inhibition extends lifespan in mammals, but mechanistic overlap with metformin, an antidiabetic drug, suggests that it might. Metformin increases insulin sensitivity by reducing blood glucose, and might increase lifespan by activating AMPK, thereby inhibiting the TOR pathway and IGF-1 signalling[65]. Glucose metabolism is a principal target of pro-longevity interventions, on the basis of the evidence that inhibiting energy pathways might mimic dietary restriction or dampen IIS. An example is 2-deoxy-D-glucose (2DG), which produces beneficial effects reminiscent of dietary restriction[8]. It is not yet known if 2DG increases lifespan, but it is at present in phase 1 clinical trial to treat solid tumours, which depend on glycolysis for survival and are sensitive to 2DG.

Natural antioxidants, such as vitamins C and E and β-carotene, have been extensively tested for their potential to extend lifespan in mice, but with no significant effects[66]. In humans, high dietary antioxidant intake correlates with decreased disease risk, but clinical trials of vitamin E and β-carotene supplementation failed to improve disease outcome[67]. Synthetic ROS scavengers have been tested in *C. elegans*[68,69] and mice[70], with conflicting results.

Free radical spin traps, such as phenyl tert-butyl-nitrone (PBN), block or reverse damage associated with a variety of disease states in animal models[71]. Thus far, however, a PBN derivative has not shown promise for extending longevity in mice[10].

Certain AGE crosslinks can be cleaved by chemical agents. At present, known crosslink breakers (for example, the thiazolium halide alagebrium) are only partially effective because they break only a small subset of crosslinked structures[72]. Nonetheless, alagebrium alleviated cardiovascular stiffening in monkeys[73] and is at present in advanced testing in humans, primarily for cardiovascular and diabetes-associated renal disease.

Rapid advances in stem cell research makes cell replacement therapy a promising approach to regenerate intrinsically aged, functionally declining tissues. Because endogenous stem cells show functional declines with age, a logical option is to differentiate pluripotent cells—obtained from embryos or generated by reprogramming of somatic cells—into specialized cells *ex vivo*. Before this approach becomes a realistic option, we must understand much more about how adult stem cells contribute to tissue maintenance during ageing, how ageing alters the stem cell microenvironment or niche, and how well the capacity of stem cells to regenerate functional tissues is maintained during differentiation and expansion in culture[74].

reducing mitochondrial function and its toxic by-products. In mammals, similar downregulation might be lethal or cause serious disease.

Translation to humans

Although disrupting conserved pro-ageing pathways identified in model organisms seems a realistic starting point for human lifespan extension, we first must determine whether these pathways modulate ageing in our own species. An initial approach is to identify associations between polymorphisms in or around conserved genes and human longevity. Extreme human longevity is genetically controlled, as indicated by the higher chance of siblings of centenarians to survive more than 100 years and moderate familial clustering of extreme longevity[26]. Thus far, however, linkage analyses are inconclusive, possibly because studies were underpowered or because of admixture in control populations[27].

Attempts to associate candidate genes with extreme human longevity have mainly identified variants in lipoprotein metabolism genes as overrepresented in centenarians[28]. Furthermore, variants in *FOXO1* and *FOXO3* genes segregated with survival to age 85 and older[29], and, in females, gene variants that reduce insulin/IGF-I signalling are associated with long survival[30]. Recently, heterozygous mutations in the IGF-IR, which markedly reduced IGF-IR activity, were found overrepresented in centenarians[31].

Although these results are promising, more research is needed to confirm that humans and model organisms use similar longevity-modulating pathways. Even if these pathways are conserved in *Homo sapiens*, their natural variation evidently does not extend lifespan as much as laboratory-generated mutations in model organisms, notably nematodes. It is possible that organismal complexity will limit how much lifespan extension can be achieved by manipulating metabolic pathways; or there could be other layers of control or pathways that are yet to be discovered in complex animals. In predicting lifespan extension in humans, it is important to remember that these are crucial questions and their answers are unknown.

Life extension in model organisms may be an artefact to some extent. None of the laboratory animals considered 'wild type' has the genetic diversity of true wild strains, nor is the laboratory a natural habitat. For example, dietary restriction does not substantially increase longevity in some wild mice. Thus, laboratory breeding might select for a robust dietary restriction response[32]. Two longitudinal dietary restriction studies in rhesus monkeys were initiated in the late 1980s. Interim results suggest that dietary restriction improves health (for example, less body fat, higher insulin sensitivity and favourable circulating lipids), but there is no evidence yet that dietary restriction increases lifespan to the extent that it does in laboratory rodents[33,34]. Moreover, in monkeys (and by extension, humans) some benefits of dietary restriction, such as low IGF-I levels, may decrease cancer risk, but also increase the risk of osteoporotic fractures[35]. Thus, it might be necessary to reduce IGF-I signalling during early adulthood to prevent cancer, but increase it at older ages to prevent non-cancerous diseases[36].

Can we expect interventions that target the human IIS pathway, even with proper spatio-temporal regulation, to extend lifespan to the extent that they do in simple models? Among the pro-longevity effects of dampening IGF-I signalling is upregulated stress resistance. Notably, stress response is generally superior in cells from long-lived compared to short-lived species[37]. In short-lived species, there is evidently sufficient opportunity for enhancing protective mechanisms. However, in long-lived species, there may be fewer such opportunities[38].

Furthermore, human physiology obviously differs from that of yeast, nematodes or flies. Perhaps less obvious are the differences between humans and mice. We should keep in mind that many anti-cancer therapies are successful in mice but fail in humans. Moreover, side or off-target effects of drugs that affect complex physiological pathways are already a problem. For example, cholesteryl ester transfer protein inhibitors, developed to increase high-density lipoprotein cholesterol, did not decrease but increased the risk of heart disease[39].

Before we can rationally evaluate the potential impact of interventions to increase human lifespan substantially, we will need to understand the primary causes of ageing, which leads to the important question of why and how we age.

Evolutionary logic of ageing

According to Dobzhansky[40], "nothing in biology makes sense except in the light of evolution"; so it is in the biology of ageing. Most scientists now accept that ageing results from the greater weight placed by natural selection on early survival and reproduction than on vigour at later ages. This age-related decline in the force of natural selection, first articulated by Medawar[41], is due to high mortality caused by extrinsic hazards in natural environments, resulting in a relative scarcity of older individuals. When these hazards make survival to old age rare, natural selection favours gene variants that

promote early growth and reproduction. In less hazardous environments, survival increases and gene variants that promote somatic maintenance can propagate. Hence, species-specific lifespan is determined by a trade-off between somatic maintenance and early growth or reproduction (Fig. 2)[42]. For example, genes that ensure a powerful immune response to infection promote early life survival, but later contribute to inflammation, a major age-related phenotype and risk for developing many diseases[43].

Obviously, huge differences in longevity can arise as a result of evolution—consider the longevity difference between nematodes (weeks) and mammals (years), or even between mice (~3 yr) and humans (~100 yr). Were these differences achieved evolutionarily by discarding pro-ageing pathways, or by creating new longevity assurance pathways? The notable conservation among known longevity-modulating pathways, and similarities between organisms such as mice and humans in genomic structure and organization, argue against this possibility. Of course, unique non-conserved pathways may yet be discovered. It is more probable that significant longevity was achieved by subtle changes in many genes over the course of evolution, not by single mutations with large effects, which often increase lifespan at a cost to reproduction or survival under stress[45]. If so, interventions that target a single mammalian gene or even a single pathway may not increase longevity to the extent achieved by natural selection. This should not discourage the search for pharmacological interventions, but rather underscores how the shallowness of our knowledge about comparative evolutionary mechanisms can severely hamper efforts in this area.

Despite a general consensus regarding the evolutionary basis of why we age, we still know little about the primary causes of ageing and its relation to disease, which is generally the cause of death.

The ageing phenotype and the relationship to disease

Emphasis on lifespan can distract from understanding ageing itself. In nematodes and flies, we know much about genes that determine lifespan, but little about how these animals die. This is due to the complexity of ageing phenotypes[42] and our limited ability to define phenotype, in contrast to the relative ease of defining genotype.

Table 1 lists some of the best-known ageing phenotypes in humans, which are increasingly evident in laboratory models as they are scrutinized. There are remarkable similarities among species, but there are also marked differences. For example, amyloid plaques in the brain and atherosclerotic plaques in blood vessels are hallmarks of human ageing, but are virtually lacking in mice. Even examination of shared phenotypes can uncover differences. For example, kyphosis (spinal curvature) is caused by osteoporosis in humans, but can have other causes, such as growth plate abnormalities, in mice. Importantly, ageing phenotypes—from hair greying to cancer susceptibility—vary among individual humans, and among inbred mouse strains.

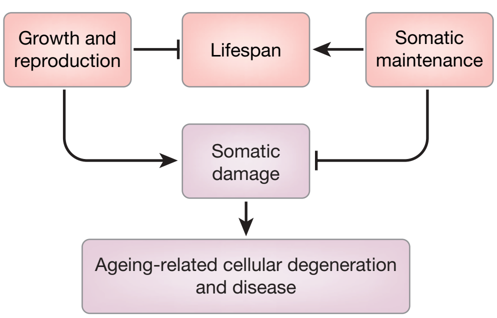


Figure 2 | Balancing somatic maintenance with growth and reproduction may determine lifespan. According to the 'disposable soma theory'[63], organisms must compromise between energy allocation to growth and reproduction or somatic maintenance and repair.

A prominent age-related phenotype in humans and mice, absent in nematodes and flies, is cancer. Cancer arises from renewable somatic tissues, which are largely lacking in invertebrates. Cancer is sometimes considered as the opposite of ageing because it entails more vigorous growth. Indeed, cellular senescence, the irreversible cessation of growth, was once considered a model for ageing *in vivo*, but is now known to be a stress and tumour suppressive response[19]. Senescent cells increase with age in mice, non-human primates and humans, but comprise only a fraction of cells in renewable tissues[19]. Cellular senescence may be another evolutionary trade-off, as it suppresses cancer at early ages, but may promote ageing by exhausting stem cells or altering their niches. Senescent cells secrete inflammatory cytokines and other molecules that alter tissue microenvironments, and could stimulate the growth of cells that harbour preneoplastic mutations[46]. On the other hand, increased cellular senescence (and decreased proliferative potential) could also explain the decrease in cancer incidence rate at very old age, that is, after age 80 (ref. 47). Hence, cellular senescence can act both as a carcinogen and as an anti-carcinogen.

Little is known about ageing phenotypes and the causes of death in *C. elegans*. The nervous system is remarkably preserved, but ageing nematodes show slower movement, lower pharyngeal pumping rates (due to muscle deterioration resembling human sarcopenia) and increased lipofuscin[48]. Notably, there is extensive variability in age-related degeneration among genetically identical animals and cells of the same type within an individual[48]. This finding emphasizes the potentially important role of stochastic events in ageing, a point to which we will return.

Also in *Drosophila*, little is known about ageing or the causes of death. Nonetheless, this organism is emerging as a powerful model for human age-related diseases, such as Parkinson's disease[49]. Ageing flies also develop sarcopenia and accumulate lipofuscin, so these traits may be universal ageing phenotypes[50]. Flies also show signs of cognitive dysfunction (increased time to accomplish a task), sharing this phenotype with mice and humans (Table 1).

Should we distinguish between ageing and disease? The answer to this question, which is still debated, depends on the disease and how its mechanism relates to 'intrinsic ageing'; that is, ageing-related changes that are not determined primarily by external factors or genetic predisposition. Early-onset diseases, such as sickle cell anaemia, caused by a heritable β-globin gene mutation, can result at young ages in vascular constriction and increased risk of infection, also common in older people. But because these phenotypes occur within the realm of natural selection, sickle cell anaemia is not an ageing-related disease as its causes have little to do with ageing. Such a mechanistic distinction is much more difficult for late-onset diseases. Many would distinguish potentially fatal vascular degeneration from benign greying of hair. However, both phenotypes could have the same cause: intrinsic ageing. On the other hand, different mechanisms might produce the same disease-related phenotype at old age. For example, intrinsic endothelial cell ageing might contribute to atherosclerosis, as do mutations or polymorphisms in genes encoding the low-density lipoprotein receptor or ApoB. Statins can lower cholesterol and suppress atherosclerosis in individuals with high-risk low-density lipoprotein receptor or ApoB alleles, but cannot prevent intrinsic endothelial cell ageing.

Diseases are the main causes of death in elderly humans. Arteriosclerosis, diabetes, dementia, osteoporosis, osteoarthritis and cancer are particularly prominent pathologies, and some account for much of mortality at old age. Among the old who escape these diseases, the cause of death is often unknown. However, because interactions among ageing phenotypes are complex, natural death may ultimately be traced to disease, even if occult. For example, subtle tissue atrophies, neuropathies or microvascular leakage may underlie the deaths of old people subjected to stress. It is not clear whether successful intervention in overt disease will ameliorate intrinsic ageing, and thus significantly extend human lifespan.

Table 1 | Conserved ageing phenotypes

Phenotype	*H. sapiens*	*M. musculus*	*D. melanogaster*	*C. elegans*
Decreased cardiac function	Yes	Yes	Yes	NA
Apoptosis, senescence (somatic cells)	Yes	Yes	Yes	?
Cancer, hyperplasia	Yes	Yes	No	No
Genome instability	Yes	Yes	Yes	Yes
Macromolecular aggregates	Yes	Yes	Yes	Yes
Reduced memory and learning	Yes	Yes	Yes	NA
Decline in GH, DHEA, testosterone, IGF	Yes	Yes	?	?
Increase in gonadotropins, insulin	Yes	Yes	?	?
Decreased thyroid function	Yes	Yes	NA	NA
Decrease in innate immunity	Yes	Yes	Yes	Yes
Increase in inflammation	Yes	Yes	No	No
Skin/cuticle morphology changes	Yes	Yes	?	Yes
Decreased mitochondrial function	Yes	Yes	Yes	Yes
Sarcopenia	Yes	Yes	Yes	Yes
Osteoporosis	Yes	Yes	NA	NA
Abnormal sleep/rest patterns	Yes	Yes	Yes	?
Decrease in vision	Yes	Yes	?	NA
Demyelination	Yes	Yes	?	No
Decreased fitness	Yes	Yes	Yes	Yes
Arteriosclerosis	Yes	No	NA	NA
Changes in fat*	Yes	Yes	?	?

* Although changes in fat content and distribution have been reported for long-lived invertebrate mutants, at present there are no data on fat-related changes during normal ageing in these organisms. GH, growth hormone; DHEA, dehydroandrosterone; NA, not applicable.

Ageing is influenced by genetic and environmental factors that may be unrelated to each other or to intrinsic ageing. Irrespective of possible intrinsic ageing mechanisms, alleles that promote ageing will penetrate the germ line as long as their adverse effects manifest late enough, creating a diversity of genetic risk factors. Even among inbred individuals, for example, monozygotic human twins, genetic diversity occurs in somatic cells by mutation and epimutation at very early ages[51,52]. Similarly, environmental or lifestyle factors (for example, sunlight or smoking) can accelerate intrinsic ageing in specific tissues. It is conceivable that individuals of extreme longevity (that is, 100 years and older) are primarily those who managed to escape these genetic and environmental risks. This possibility is supported by the decelerated mortality rate seen at old age in invertebrate and human populations, indicating the survival of increasingly less frail individuals[53]. A question remains as to whether these survivors that escape the genetic and environmental risks that normally eliminate individuals through disease succumb to intrinsic ageing. Is there really an intrinsic ageing mechanism(s) to which eventually every cell or tissue falls prey? If so, what is its basis?

Intrinsic ageing

Ageing entails numerous functional and structural changes, many, but not all of which, adversely affect survival. A universal process of 'intrinsic ageing' might explain common ageing phenotypes among animals. One characteristic shared by all species studied thus far is the accumulation of unrepaired somatic damage. Thus, lifelong accumulation of various types of damage, along with random errors in bioinformational processes, might underlie intrinsic ageing (Fig. 3). As discussed, attenuation of such damage could explain the longevity conferred by mutations that dampen normal metabolic processes. Moreover, defence systems that keep damage in check might differ in efficacy among species, thereby dictating their lifespan[37].

Prominent causes of somatic damage include reactive oxygen species (ROS) and reducing sugars. ROS (by-products of respiration and other metabolic processes[54]) can damage and crosslink DNA, proteins and lipids. Reducing sugars react with carbohydrates and free amino groups, resulting in difficult-to-degrade advanced glycation end products (AGEs)[55]. AGEs accumulate in long-lived structural proteins, such as collagen and elastin. They increase the stiffness of blood vessels, joints and the bladder, and impair function in the kidney, heart, retina and other organs.

Interventions that remove damage might successfully counter the adverse effects of ROS and AGEs, thereby postponing ageing indefinitely[56] (see Box 1). However, macromolecular damage comes in many forms and all key lesions may not have been identified. Moreover, we do not know their relative contributions to intrinsic ageing, or how various components of the damage spectrum interact.

It might seem more efficacious to eliminate damaging molecules, rather than damage itself. However, some damaging molecules are crucial for normal cellular function. Two examples are glucose, which clearly cannot be eliminated, and ROS, which are also signalling molecules[57]. Trade-offs between beneficial and deleterious effects of damaging molecules will complicate strategies aimed at extending longevity by neutralizing them.

A similar trade-off might be cellular processes that defend us against cancer. Tissue regeneration elevates cancer risk by increasing the chance of acquiring DNA mutations or epimutations, which occur frequently in every organism as a consequence of errors during the repair or replication of a damaged template. Tumour suppressor mechanisms either eliminate cells that acquired extensive damage (apoptosis) or permanently prevent their proliferation (senescence). These responses, however, may gradually cause tissue atrophy, and

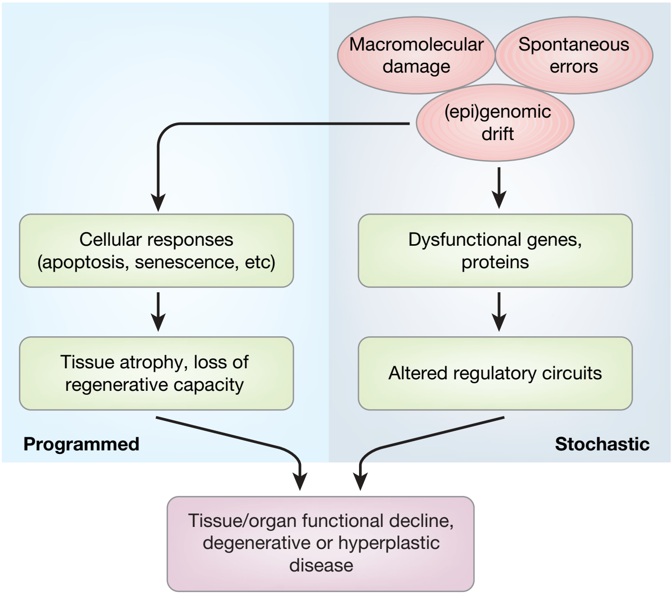


Figure 3 | The causes of intrinsic ageing. Although ultimately stochastic in nature, the proximal causes of ageing involve both programmed and random mechanisms.

therefore loss of organ function and regenerative capacity (Fig. 3)[58]. In principle, stem cell transplantation could counter the adverse effects of these damage responses.

When damage levels are not high enough to elicit an apoptotic or senescence response, a potentially more serious situation ensues that is difficult to counter: the gradual accumulation of random changes in DNA or protein, turning tissues into cellular mosaics. Such stochastic changes can reset gene regulatory loops, and randomly alter gene expression patterns on a cell by cell basis[59,60]. These changes, in aggregate, might compromise tissue function, without eliciting immediate cellular responses (Fig. 3). Stochastic gene regulatory drift would be difficult to correct and could even occur in stem cells *in vivo* or *ex vivo* during expansion for transplantation therapy.

Furthermore, developmental pathways that are essential for early life fitness or reproduction might be deleterious in adult tissues—for example, pathways that drive ductal morphogenesis in the developing or pregnant mammary gland might gradually promote ductal hyperplasia in the adult gland, predisposing it to cancer.

Future prospects

There is abundant evidence for mortality decline and increased lifespan throughout the developed world[1]. Before 1970, this probably reflected improvements in the food supply and sanitation, and achievements in medicine, notably vaccination and antibiotics.

After 1970, the mortality decline probably reflects preventive medicine, lifestyle changes, routine use of anti-hypertensive and other drugs, and so on. It is impossible to predict, at this stage, whether increasingly sophisticated interventions will negate most or all ageing phenotypes. However, there are reasons for caution.

First, pharmacological intervention on the basis of pathways identified in model organisms may be an illusion because gains in longevity achieved in these organisms seem to decline with organismal complexity or depend on idiosyncratic physiology. Furthermore, lifespan in some organisms may be less plastic than in others. In addition, cancer poses a challenge to longevity that is distinct from age-related degeneration, and may be suppressed by mechanisms that are also pro-ageing. Third, there are still enormous gaps in our knowledge about how metabolic pathways operate and interact; serious side effects may constrain the effectiveness of pharmacological interventions.

Repair of macromolecular damage may prove more promising, especially in unison with improved anti-cancer therapies. However, it is not clear that all toxic lesions have been identified, or whether practical strategies exist to eliminate them. For example, it would be impossible to counter (epi)genomic drift pharmacologically, and transplanted organs and cells are also subject to loss of (epi)genomic integrity. Moreover, we do not know if macromolecular damage is the sole cause of ageing. Even in simple organisms, it is clear that longevity-modulating pathways entail exquisitely balanced interactions, regulated by numerous genetic elements[61]. And the large number of genomic transactions makes errors—many that are irreversible—inevitable. Indeed, there is evidence that ageing entails a gradual drift towards more random patterns of gene expression[59], which might cause organ/tissue failure that cannot be undone by pharmacological or biological intervention.

In theory, interventions could be designed to alter the orchestrated networks of cell–cell interaction to increase lifespan. This is essentially what evolution has done to produce long-lived species. The question is, can we mimic the evolutionary process to the extent that senescence becomes essentially negligible? At this stage, the answer must be that we do not know. Although there is no scientific reason for not striving to cure ageing—similar to what we profess to do for cancer and other diseases—our current understanding makes it impossible to assert that indefinite postponement is feasible. Rather, we need to use the current momentum to intensify research aimed at resolving major outstanding questions that hinder a more complete understanding of basic ageing mechanisms and their relationship to disease (Box 2). Only this will allow us to generate sophisticated, integrated strategies to increase human health and lifespan.

Box 2 | Future research

Comparative phenotyping

In view of the general agreement that any ageing intervention should primarily target the quality of late life, it is imperative to learn more about ageing phenotypes and how the human phenotypes compare to those in model animals. This goal hinges on good descriptive research, traditionally less appreciated by the scientific community than hypothesis testing, but in this case it is critically important. This type of research should include a search for biomarkers, not as long-sought but illusory predictors of individual mortality risk, but to better define and characterize the degenerative processes underlying mortality. Furthermore, progress in this area is unthinkable without the development and use of advanced, interactive database systems[75].

The ageing–disease relationship

To evaluate the effects of ageing interventions, it is important to understand how the pathogenesis of age-related diseases (for example, cancer, diabetes, cardiovascular and neurodegenerative disorders), which account for most of late life mortality, relates to basic molecular processes of ageing. Although we need to abandon the concept that ageing is a genetically programmed process and diseases of ageing essentially separate entities, it is conceivable that many disease risk factors are independent of ageing *per se*, which may have only one or a few, rather than a multiplicity, of causes. It is important to begin deciphering how disease processes intersect with basic ageing cause(s).

A rational basis for interventions

Before developing interventions for humans, it is critically important to have a strong rationale for both the intervention and target—this rationale should derive from many independent lines of evidence, including mechanistic, animal model and human-based studies. In selecting targets, strong consideration should be given to potential side effects. This mission will necessarily require multidisciplinary teams of clinicians and basic scientists working in cooperation with the pharmaceutical industry and regulatory agencies.

Unravel the causes of ageing

All interventions will fail to increase longevity substantially if the ultimate cause(s) of ageing does not lend itself to available treatments. For example, if ageing involves a gradual drift away from regulatory constraints, with organs and tissues developing into cellular mosaics, only *in situ* tissue replacement may offer a long-term solution. Even if ultimately ageing has the singular cause of a gradual, lifelong accumulation of molecular and cellular damage, the large diversity in types of damage, their complex interactions and ultimate consequences require an integrative, systems-biology approach[42].

1. Kinsella, K. G. Future longevity-demographic concerns and consequences. *J. Am. Geriatr. Soc.* **53**, S299–S303 (2005).
2. Olshansky, S. J., Carnes, B. A. & Cassel, C. In search of Methuselah: estimating the upper limits to human longevity. *Science* **250**, 634–640 (1990).
3. Klass, M. R. A method for the isolation of longevity mutants in the nematode *Caenorhabditis elegans* and initial results. *Mech. Ageing Dev.* **22**, 279–286 (1983).
4. Friedman, D. B. & Johnson, T. E. A mutation in the *age-1* gene in *Caenorhabditis elegans* lengthens life and reduces hermaphrodite fertility. *Genetics* **118**, 75–86 (1988).
5. Kenyon, C. The plasticity of aging: insights from long-lived mutants. *Cell* **120**, 449–460 (2005).
6. Masoro, E. J. Overview of caloric restriction and ageing. *Mech. Ageing Dev.* **126**, 913–922 (2005).
7. Gems, D., Pletcher, S. & Partridge, L. Interpreting interactions between treatments that slow aging. *Aging Cell* **1**, 1–9 (2002).
8. Ingram, D. K. *et al.* Calorie restriction mimetics: an emerging research field. *Aging Cell* **5**, 97–108 (2006).
9. Chen, D. & Guarente, L. SIR2: a potential target for calorie restriction mimetics. *Trends Mol. Med.* **13**, 64–71 (2007).
10. Miller, R. A. *et al.* An aging interventions testing program: study design and interim report. *Aging Cell* **6**, 565–575 (2007).
11. Botstein, D., Chervitz, S. A. & Cherry, J. M. Yeast as a model organism. *Science* **277**, 1259–1260 (1997).
12. Santini, M. P. *et al.* Enhancing repair of the mammalian heart. *Circ. Res.* **100**, 1732–1740 (2007).
13. Pelosi, L. *et al.* Local expression of IGF-1 accelerates muscle regeneration by rapidly modulating inflammatory cytokines and chemokines. *FASEB J.* **21**, 1393–1402 (2007).

14. Bluher, M., Kahn, B. B. & Kahn, C. R. Extended longevity in mice lacking the insulin receptor in adipose tissue. *Science* **299**, 572–574 (2003).
15. Selman, C. *et al.* Evidence for lifespan extension and delayed age-related biomarkers in insulin receptor substrate 1 null mice. *FASEB J.* **22**, 807–818 (2008).
16. Tran, H. *et al.* DNA repair pathway stimulated by the forkhead transcription factor FOXO3a through the Gadd45 protein. *Science* **296**, 530–534 (2002).
17. Giannakou, M. E. & Partridge, L. The interaction between FOXO and SIRT1: tipping the balance towards survival. *Trends Cell Biol.* **14**, 408–412 (2004).
18. Smith, J. Human Sir2 and the 'silencing' of p53 activity. *Trends Cell Biol.* **12**, 404–406 (2002).
19. Campisi, J. & d'Adda di Fagagna, F. Cellular senescence: when bad things happen to good cells. *Nature Rev. Mol. Cell Biol.* **8**, 729–740 (2007).
20. Ayyadevara, S., Alla, R., Thaden, J. J. & Shmookler Reis, R. J. Remarkable longevity and stress resistance of nematode PI3K-null mutants. *Aging Cell* **7**, 13–22 (2007).
21. Broughton, S. J. *et al.* Longer lifespan, altered metabolism, and stress resistance in *Drosophila* from ablation of cells making insulin-like ligands. *Proc. Natl Acad. Sci. USA* **102**, 3105–3110 (2005).
22. Bartke, A. Minireview: role of the growth hormone/insulin-like growth factor system in mammalian aging. *Endocrinology* **146**, 3718–3723 (2005).
23. Clancy, D. J., Gems, D., Hafen, E., Leevers, S. J. & Partridge, L. Dietary restriction in long-lived dwarf flies. *Science* **296**, 319 (2002).
24. Spencer, C. C., Howell, C. E., Wright, A. R. & Promislow, D. E. Testing an 'aging gene' in long-lived *Drosophila* strains: increased longevity depends on sex and genetic background. *Aging Cell* **2**, 123–130 (2003).
25. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33**, 40–48 (2003).
26. Perls, T. *et al.* Survival of parents and siblings of supercentenarians. *J. Gerontol. A Biol. Sci. Med. Sci.* **62**, 1028–1034 (2007).
27. Beekman, M. *et al.* Chromosome 4q25, microsomal transfer protein gene, and human longevity: novel data and a meta-analysis of association studies. *J. Gerontol. A Biol. Sci. Med. Sci.* **61**, 355–362 (2006).
28. Barzilai, N. *et al.* Unique lipoprotein phenotype and genotype associated with exceptional longevity. *J. Am. Med. Assoc.* **290**, 2030–2040 (2003).
29. Kuningas, M. *et al.* Haplotypes in the human *Foxo1a* and *Foxo3a* genes; impact on disease and mortality at old age. *Eur. J. Hum. Genet.* **15**, 294–301 (2007).
30. Capri, M. *et al.* The genetics of human longevity. *Ann. NY Acad. Sci.* **1067**, 252–263 (2006).
31. Suh, Y. *et al.* Functionally significant insulin-like growth factor I receptor mutations in centenarians. *Proc. Natl Acad. Sci. USA* **105**, 3438–3442 (2008).
32. Harper, J. M., Leathers, C. W. & Austad, S. N. Does caloric restriction extend life in wild mice? *Aging Cell* **5**, 441–449 (2006).
33. Mattison, J. A., Lane, M. A., Roth, G. S. & Ingram, D. K. Calorie restriction in rhesus monkeys. *Exp. Gerontol.* **38**, 35–46 (2003).
34. Weindruch, R. Will dietary restriction work in primates? *Biogerontology* **7**, 169–171 (2006).
35. Dirks, A. J. & Leeuwenburgh, C. Caloric restriction in humans: potential pitfalls and health concerns. *Mech. Ageing Dev.* **127**, 1–7 (2006).
36. Yang, J., Anzo, M. & Cohen, P. Control of aging and longevity by IGF-I signaling. *Exp. Gerontol.* **40**, 867–872 (2005).
37. Kapahi, P., Boulton, M. E. & Kirkwood, T. B. Positive correlation between mammalian lifespan and cellular resistance to stress. *Free Radic. Biol. Med.* **26**, 495–500 (1999).
38. *Accuracy of Molecular Processes* (eds Kirkwood, T. B., Rosenberger, R. F., Galas, D. J.) (Chapman & Hall, (1986).
39. Suckling, K. The continuing complexities of high-density lipoprotein metabolism in drug discovery and development. *Expert Opin. Ther. Targets* **11**, 1133–1136 (2007).
40. Dobzhansky, T. Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* **35**, 125–129 (1973).
41. Medawar, P. B. *An Unsolved Problem in Biology* (H. K. Lewis, 1952).
42. Kirkwood, T. B. A systematic look at an old problem. *Nature* **451**, 644–647 (2008).
43. Finch, C. E. & Crimmins, E. M. Inflammatory exposure and historical changes in human life-spans. *Science* **305**, 1736–1739 (2004).
44. Kirkwood, T. B. Understanding the odd science of aging. *Cell* **120**, 437–447 (2005).
45. Jenkins, N. L., McColl, G. & Lithgow, G. J. Fitness cost of extended lifespan in *Caenorhabditis elegans. Proc. Biol. Sci.* **271**, 2523–2526 (2004).
46. Krtolica, A., Parrinello, S., Lockett, S., Desprez, P. Y. & Campisi, J. Senescent fibroblasts promote epithelial cell growth and tumorigenesis: a link between cancer and aging. *Proc. Natl Acad. Sci. USA* **98**, 12072–12077 (2001).
47. Harding, C., Pompei, F., Lee, E. E. & Wilson, R. Cancer suppression at old age. *Cancer Res.* **68**, 4465–4478 (2008).
48. Herndon, L. A. *et al.* Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans. Nature* **419**, 808–814 (2002).
49. Wang, C. *et al. Drosophila* overexpressing parkin R275W mutant exhibits dopaminergic neuron degeneration and mitochondrial abnormalities. *J. Neurosci.* **27**, 8563–8570 (2007).
50. Grotewiel, M. S., Martin, I., Bhandari, P. & Cook-Wiens, E. Functional senescence in *Drosophila melanogaster. Ageing Res. Rev.* **4**, 372–397 (2005).
51. Bruder, C. E. *et al.* Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am. J. Hum. Genet.* **82**, 763–771 (2008).
52. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* **102**, 10604–10609 (2005).
53. Vaupel, J. W. *et al.* Biodemographic trajectories of longevity. *Science* **280**, 855–860 (1998).
54. Muller, F. L., Lustgarten, M. S., Jang, Y., Richardson, A. & Van Remmen, H. Trends in oxidative aging theories. *Free Radic. Biol. Med.* **43**, 477–503 (2007).
55. Ulrich, P. & Cerami, A. Protein glycation, diabetes, and aging. *Recent Prog. Horm. Res.* **56**, 1–21 (2001).
56. de Grey, A. D. *et al.* Time to talk SENS: critiquing the immutability of human aging. *Ann. NY Acad. Sci.* **959**, 452–62 (2002).
57. Finkel, T. Oxidant signals and oxidative stress. *Curr. Opin. Cell Biol.* **15**, 247–254 (2003).
58. Sharpless, N. E. & DePinho, R. A. How stem cells age and why this makes us grow old. *Nature Rev. Mol. Cell Biol.* **8**, 703–713 (2007).
59. Bahar, R. *et al.* Increased cell-to-cell variation in gene expression in aging mouse heart. *Nature* **441**, 1011–1014 (2006).
60. Vijg, J. *Aging of the Genome* (Oxford Univ. Press, 2007).
61. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
62. DiMauro, S. & Schon, E. A. Mitochondrial respiratory-chain diseases. *N. Engl. J. Med.* **348**, 2656–2668 (2003).
63. Kirkwood, T. B. Evolution of ageing. *Nature* **270**, 301–304 (1977).
64. Baur, J. A. *et al.* Resveratrol improves health and survival of mice on a high-calorie diet. *Nature* **444**, 337–342 (2006).
65. McCarty, M. F. Chronic activation of AMP-activated kinase as a strategy for slowing aging. *Med. Hypotheses* **63**, 334–339 (2004).
66. Meydani, M. *et al.* The effect of long-term dietary supplementation with antioxidants. *Ann. NY Acad. Sci.* **854**, 352–360 (1998).
67. Frei, B. Efficacy of dietary antioxidants to prevent oxidative damage and inhibit chronic disease. *J. Nutr.* **134**, 3196S–3198S (2004).
68. Melov, S. *et al.* Extension of life-span with superoxide dismutase/catalase mimetics. *Science* **289**, 1567–1569 (2000).
69. Keaney, M. & Gems, D. No increase in lifespan in *Caenorhabditis elegans* upon treatment with the superoxide dismutase mimetic EUK-8. *Free Radic. Biol. Med.* **34**, 277–282 (2003).
70. Morten, K. J., Ackrell, B. A. & Melov, S. Mitochondrial reactive oxygen species in mice lacking superoxide dismutase 2: attenuation via antioxidant treatment. *J. Biol. Chem.* **281**, 3354–3359 (2006).
71. Floyd, R. A. Nitrones as therapeutics in age-related diseases. *Aging Cell* **5**, 51–57 (2006).
72. Furber, J. D. Extracellular glycation crosslinks: prospects for removal. *Rejuvenation Res.* **9**, 274–278 (2006).
73. Vaitkevicius, P. V. *et al.* A cross-link breaker has sustained effects on arterial and ventricular properties in older rhesus monkeys. *Proc. Natl Acad. Sci. USA* **98**, 1171–1175 (2001).
74. Rando, T. A. Stem cells, ageing and the quest for immortality. *Nature* **441**, 1080–1086 (2006).
75. Calder, R. B. *et al.* MPHASYS: a mouse phenotype analysis system. *BMC Bioinformatics* **8**, 183 (2007).

Acknowledgements We thank N. Barzilai, A. de Grey, G. Lithgow and M. Gough for comments on the manuscript and P. Kapahi, R. Shmookler Reis, L. Balducci and Y. Suh for discussions. The authors' work is supported by the US National Institutes of Health and Ellison Medical Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to J.C. (JCampisi@lbl.gov) or J.V. (jvijg@aecom.yu.edu).

Vol 454|28 August 2008|doi:10.1038/nature07243

ARTICLES

How Cooper pairs vanish approaching the Mott insulator in $Bi_2Sr_2CaCu_2O_{8+\delta}$

Y. Kohsaka[1,2], C. Taylor[1], P. Wahl[1], A. Schmidt[1], Jhinhwan Lee[1], K. Fujita[1,3], J. W. Alldredge[1,4], K. McElroy[4], Jinho Lee[1,5,6], H. Eisaki[7], S. Uchida[3], D.-H. Lee[8] & J. C. Davis[1,6]

The antiferromagnetic ground state of copper oxide Mott insulators is achieved by localizing an electron at each copper atom in real space (r-space). Removing a small fraction of these electrons (hole doping) transforms this system into a superconducting fluid of delocalized Cooper pairs in momentum space (k-space). During this transformation, two distinctive classes of electronic excitations appear. At high energies, the mysterious 'pseudogap' excitations are found, whereas, at lower energies, Bogoliubov quasi-particles—the excitations resulting from the breaking of Cooper pairs—should exist. To explore this transformation, and to identify the two excitation types, we have imaged the electronic structure of $Bi_2Sr_2CaCu_2O_{8+\delta}$ in r-space and k-space simultaneously. We find that although the low-energy excitations are indeed Bogoliubov quasi-particles, they occupy only a restricted region of k-space that shrinks rapidly with diminishing hole density. Concomitantly, spectral weight is transferred to higher energy r-space states that lack the characteristics of excitations from delocalized Cooper pairs. Instead, these states break translational and rotational symmetries locally at the atomic scale in an energy-independent way. We demonstrate that these unusual r-space excitations are, in fact, the pseudogap states. Thus, as the Mott insulating state is approached by decreasing the hole density, the delocalized Cooper pairs vanish from k-space, to be replaced by locally translational- and rotational-symmetry-breaking pseudogap states in r-space.

Evolution from CuO_2 superconductor to Mott insulator

Theoretical studies of the copper oxide electronic structure using the Hubbard model[1–13] suggest a complex phenomenology for the evolution from the localized **r**-space states of the Mott insulator to the delocalized **k**-space Cooper pairs of the superconductor. The unique capability of quasi-particle interference (QPI) imaging[14] to determine the electronic structure simultaneously in **r**-space and **k**-space makes it ideal for studying such effects. Here we use superconducting QPI[15–19] to study the evolution of electronic structure as hole density p tends towards zero in $Bi_2Sr_2CaCu_2O_{8+\delta}$. We find that the Bogoliubov quasi-particle (BQP) excitations are confined to a specific locus in **k**-space called the 'Bogoliubov arc'. The end points of this arc, where BQP interference disappears, lie near the diagonal lines connecting $\mathbf{k}=(0,\pm\pi/a_0)$ and $\mathbf{k}=(\pm\pi/a_0,0)$ and occur at a weakly doping-dependent 'extinction energy' $E=\Delta_0$ (a_0 is the nearest neighbour Cu–Cu distance). This Bogoliubov arc, which represents regions of **k**-space supporting delocalized Cooper pairs, shrinks rapidly as p decreases.

For $E>\Delta_0$, the excited states appear dramatically different. They do not have the BQP interference phenomena expected of excitations from delocalized Cooper pairs. Instead, they appear to be quasi-localized in **r**-space. Also, energy-resolved tunnelling asymmetry[20], when corrected appropriately for nanoscale electronic disorder, reveals that the intensity of these **r**-space characteristics is most pronounced at the pseudogap energy. Thus, the low-p pseudogap excitations locally break translational symmetry, and reduce the 90° rotational (C_4) symmetry of each four-Cu-atom plaquette to a 180° rotational symmetry (C_2) in Cu–O–Cu bond-centred patterns[20].

Two classes of CuO_2 excited states at low hole density

The hole-doped CuO_2 plane has a complex electronic phase diagram[21]. Its ground states include antiferromagnetism for $p<2\%$–5%, d-wave superconductivity for 5%–$10\%<p<25\%$, and a metallic state for $p>25\%$. Several poorly understood regions occur at finite temperatures in the non-superconducting state, most significantly the pseudogap regime[21]. Its spectrum of electronic excitations reveals the pseudogap phenomenology most transparently. Two distinct energy scales, associated with two distinct types of excited state dynamics, can be detected by numerous spectroscopies[22]. The two excitation energies diverge with diminishing p as shown, for example, in Fig. 1a (reproduced from ref. 22, the nomenclature of which we follow in this introduction).

Single-particle tunnelling spectroscopy[23] detects a particle–hole-symmetric excitation energy E_{PG} that is indistinguishable in magnitude in the pseudogap and superconducting states[23] and which increases with decreasing p. A typical evolution of such spectra with decreasing p is shown in Fig. 1b. Here the two classes of excitations are respectively spatially heterogeneous[24–26] excitations centred on E_{PG} and spatially homogeneous[16,19,24–26] excitations surrounding $E=0$; they exhibit increasing energy segregation as $p\rightarrow 0$. Andreev–St James tunnelling, which is sensitive to the phase-coherent superconducting state, also reveals two diverging excitation energy scales as $p\rightarrow 0$: the first is the familiar high-energy scale E_{PG} but the second, lower energy, scale E_{SC} is viewed as the maximum energy of Cooper-pair binding[27]. Optical transient-grating spectroscopy[28] shows that the higher energy excitations near E_{PG} propagate exceedingly slowly and without recombination to form Cooper pairs,

[1]LASSP, Department of Physics, Cornell University, Ithaca, New York 14853, USA. [2]RIKEN, Wako, Saitama 351-0198, Japan. [3]Department of Physics, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. [4]Department of Physics, University of Colorado, Boulder, Colorado 80309, USA. [5]School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS, UK. [6]CMPMS Department, Brookhaven National Laboratory, Upton, New York 11973, USA. [7]Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-8568, Japan. [8]Department of Physics, University of California, Berkeley, California 94720, USA.

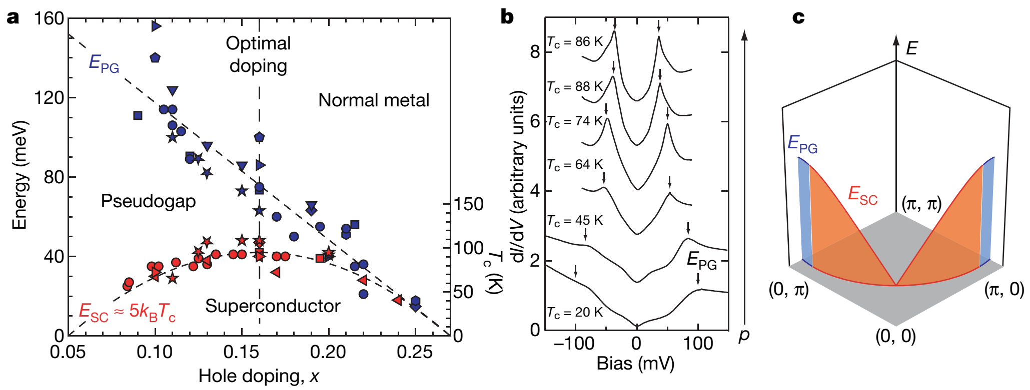


Figure 1 | Two classes of electronic excitations in copper oxides as $p \to 0$. **a**, The separation between the energy scales associated with excitations of the superconducting state (E_{SC}) and those of the pseudogap state (E_{PG}) increases as p decreases (reproduced from ref. 22). The different symbols correspond to the use of different experimental techniques. **b**, The average of all conductance spectra exhibiting the spatially averaged value of Δ_1, for each of six independently measured $Bi_2Sr_2CaCu_2O_{8+\delta}$ samples. The observed gap maximum increases along the E_{PG} trajectory as hole density is reduced[23]. Note that the peak energies $\pm E_{PG}$ are symmetric about the Fermi level (except perhaps at $p \approx 6\%$, or $T_c \approx 20$ K, where they are quite difficult to discern). **c**, The conventional view of the electronic structure in one-quarter of the CuO_2 first Brillouin zone with energy (E) plotted as a third dimension. The d-wave superconducting energy gap E_{SC}, which opens for $T < T_c$, is shown in orange, and the temperature-independent pseudogap, which opens for $T \gg T_c$, is shown in blue.

whereas lower energy excitations surrounding the $E = 0$ nodes propagate and dynamically reform Cooper pairs as expected. Raman scattering spectroscopy[29] reveals the two distinct energy scales and quasi-particle dynamics in the nodal and antinodal regions of **k**-space, again diverging as $p \to 0$.

Finally, angle-resolved photoemission spectroscopy (ARPES) plays a central role in categorizing these two excitation types. In the pseudogap regime, ARPES reveals an excitation energy E_{PG} in the antinodal region of momentum space near $\mathbf{k} \approx (\pm\pi/a_0, 0)$ and $\mathbf{k} \approx (0, \pm\pi/a_0)$ that increases as $p \to 0$. E_{PG} is independent of temperature, changing neither with greatly increased temperatures nor during the transition into the superconducting state at low temperature[22,30–34]. In contrast, the nodal region supports a 'Fermi arc'[30] of coherent excitations in the pseudogap state, upon which a momentum- and temperature-dependent gap opens only in the superconducting state[30,32–34]. A schematic representation of this situation (with E_{PG} in blue and E_{SC} in orange, following ref. 22) is shown in Fig. 1c. Notably, E_{SC} links smoothly to the pre-existing pseudogap excitation energy E_{PG} at their common point in **k**-space[30–34]. To understand the copper oxide superconductor-to-Mott insulator transition fully, it is thus critical to understand these two electronic excitation types along with their growing energetic segregation as $p \to 0$.

Numerical studies of hole-doped CuO_2 Mott insulators

The hypothesis of a 'resonating valence-bond' state[35] has motivated the theoretical exploration of whether repulsive electron–electron interactions in the CuO_2 plane could provide the explanation for high-temperature superconductivity. In such proposals, when the antiferromagnetic Mott insulator is 'quantum melted' by hole doping, singlet electron-pairing correlations develop, with the pairing energy diminishing as p increases. However, such theoretical approaches have faced a number of challenges, including how to adequately represent the anisotropic electronic structure in **k**-space[29–34] and the quasi-localized[20,24–26] **r**-space states, which become increasingly important as $p \to 0$.

Recent theoretical advances in the numerical analysis of the Hubbard model[1–9] with associated analytical approximations[10–13] are highly relevant to these issues. They yield the following results, among others: the lightly hole-doped Mott insulator state has two excitation energy scales that diverge with decreasing p (refs 1, 2, 7–9); the smaller scale represents the energy gap to excitations of delocalized Cooper pairs, whereas the larger scale is due to correlations and not a different order parameter[1–3,7–9]; a growing **k**-space anisotropy between these two types of excitation leads to an eventual break-up of the Fermi surface[1,4,5,7–11,13] as $p \to 0$; the Luttinger theorem relating the areas enclosed by zero-energy **k**-space contours to carrier density must thus be amended[1,9–13]; and the doped Mott insulator correlations produce an asymmetry of tunnelling probabilities for electron injection versus extraction[3,4,6,9,13] that diverges with decreasing p.

Simultaneous r-space and k-space electronic structure determination

To explore these predictions, the electronic structure must be determined simultaneously in **r**-space and **k**-space, as a function of decreasing p. Atomic-resolution spectroscopic-imaging scanning tunnelling microscopy in combination with BQP interference[15–19] is one of the few suitable approaches. This technique has a critical advantage in that it can detect the quantum coherence of **k**-space excitations directly from their interference patterns.

A distinctive form of QPI occurs in copper oxide superconductors because the Bogoliubov quasi-particle dispersion $E(\mathbf{k})$ has closed constant-energy contours surrounding the d-wave nodes as shown by grey lines in Fig. 2a. Maxima in the joint density of states occur at the eight tips of these closed contours, $\mathbf{k}_j(E)$, $j = 1, 2, \ldots, 8$, as shown by colour-coded symbols in Fig. 2a. For a given E, these $\mathbf{k}_j(E)$ occur at the minima (maxima) in the band structure of Bogoliubov excitations $\mathbf{k}_B(E)$ and at the normal-state Fermi surface. Elastic scattering between the $\mathbf{k}_j(E)$ is represented by the coloured arrows in Fig. 2a, b and produces **r**-space interference patterns in the local density of states $N(\mathbf{r}, E)$. The resulting energy-dispersive $N(\mathbf{r}, E)$ modulations have 16 $\pm\mathbf{q}$ pairs of wavevectors, indicated by the colour-coded markers in Fig. 2b. By inspection, the relationship between the $\mathbf{q}_i(E)$, $i = 1, 2, \ldots, 7$, and the locus of the Bogoliubov band minima $\mathbf{k}_B(E) = (k_x(E), k_y(E))$ are given by:

$$\mathbf{q}_1 = (2k_x, 0) \quad (1)$$

$$\mathbf{q}_2 = (k_x + k_y, k_y - k_x) \quad (2)$$

$$\mathbf{q}_3 = (k_x + k_y, k_y + k_x) \quad (3)$$

$$\mathbf{q}_4 = (2k_x, 2k_y) \quad (4)$$

$$\mathbf{q}_5 = (0, 2k_y) \quad (5)$$

$$\mathbf{q}_6 = (k_x - k_y, k_y + k_x) \quad (6)$$

$$\mathbf{q}_7 = (k_x - k_y, k_y - k_x) \quad (7)$$

This is referred to[15–19] as the 'octet model'. When these $\mathbf{q}_i(E)$ are measured from the Fourier transform of spatial modulations[16,19] seen in differential tunnelling conductance $dI(\mathbf{r}, V)/dV \equiv g(\mathbf{r}, V)$, the $\mathbf{k}_B(E)$ can be determined by using equations (1)–(7) with the requirement that all independent solutions be consistent. The structure of the superconductor's energy gap $\Delta(\mathbf{k})$ is then determined[15–19] directly from $\mathbf{k}_B(E)$ (see Supplementary Information, sections III and IV). Of primary significance here is that, because only the BQP states of a d-wave superconductor could exhibit such a particle–hole-symmetric set of interference wavevectors in which all dispersions are internally consistent within the octet model, the gap $\Delta(\mathbf{k})$ determined by these procedures is definitely the superconducting energy gap[16,19].

These procedures are demonstrably successful near optimal doping. In $Bi_2Sr_2CaCu_2O_{8+\delta}$, measurements from QPI of $\mathbf{k}_B(E)$ and the superconducting $\Delta(\mathbf{k})$ are consistent with ARPES[16], whereas in $Ca_{2-x}Na_xCuO_2Cl_2$, QPI yields $\mathbf{k}_B(E)$ and $\Delta(\mathbf{k})$ measurements equally well[19]. For unknown reasons, when $E <$ 3–6 meV in both systems, the modulations appear to be too weak to analyse (although with unitary scatterers these intra-nodal scattering modulations do appear). In any case, the fundamental $\mathbf{k}$-space phenomenology behind this d-wave BQP interference model[15,17,18] has also been demonstrated directly in ARPES studies[36–38]. QPI studies have been more challenging as $p \to 0$, however, because intense atomic-scale spatial fluctuations in electronic structure cause systematic errors in setting the scanning tunnelling microscope tip elevation[19,20]. Recently, though, it was shown[19] how to enhance the QPI signatures by using the ratio of differential conductances at opposite bias:

$$Z(\mathbf{r}, E = eV) \equiv \frac{g(\mathbf{r}, +V)}{g(\mathbf{r}, -V)}$$

The advantage of this procedure is that it cancels[19,20] the severe systematic errors in $g(\mathbf{r}, V)$ due to tip elevation errors, yet retains all the $\mathbf{q}_i(E)$ information (see Supplementary Information, section II).

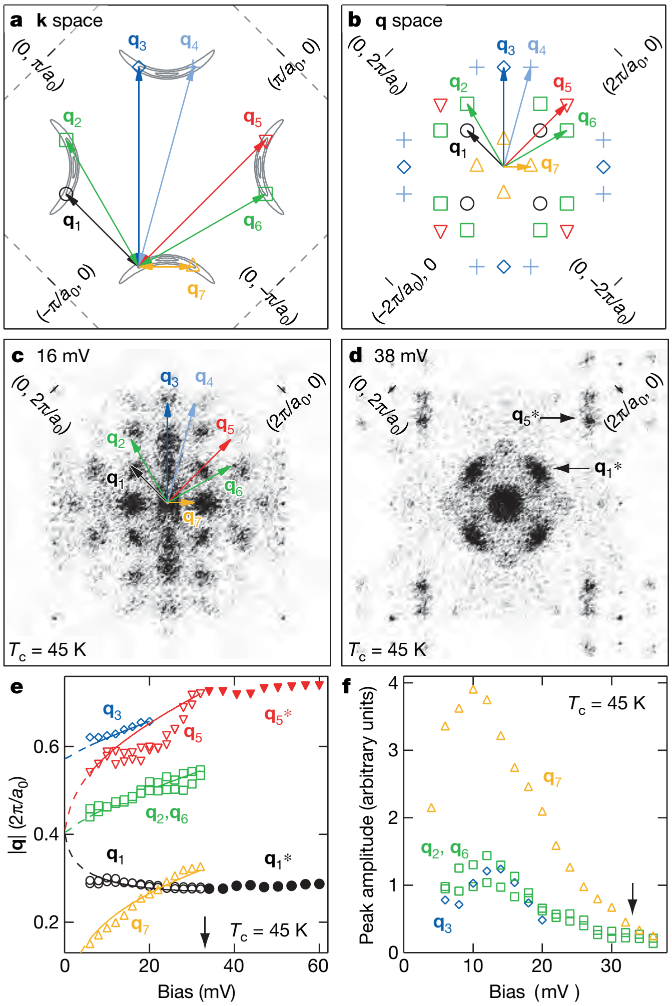


Figure 2 | BQP interference. **a**, The contours of constant energy in the superconducting state. The locations of the eight regions of maximum joint density of states are shown by coloured symbols. Quasi-particle scattering between these eight regions produces interference patterns associated with the dominant modulations in $dI(\mathbf{r}, V)/dV$. The expected wavevectors of these modulations are labelled $\mathbf{q}_i$, $i = 1, 2, \ldots, 7$, in the appropriate colour. **b**, The set of 16 pairs of wavevectors $\mathbf{q}_i(E)$ representing the octet model described in **a** constitute a highly overdetermined set in $\mathbf{q}$-space. **c**, The $\mathbf{q}$-space structure of interference patterns $Z(\mathbf{q}, E = 16\,\text{mV})$ measured on the sample with $T_c = 45$ K is consistent with the scattering patterns predicted from the octet model ideas of **a** and **b**. **d**, For comparison with **c**, we show $Z(\mathbf{r}, E = 38\,\text{mV})$, in which the interference patterns are no longer obviously associated with quasi-particle interference patterns as in **b** and, furthermore, retain this unchanging $\mathbf{q}$-space structure for a very wide range of energies $E > \Delta_0$. **e**, The magnitude of various extracted scattering vectors, labelled with the same colours and symbols as in **a** and **b** and plotted as a function of energy. Whereas the expected energy dispersion of the octet vectors $\mathbf{q}_i(E)$ is apparent for $|E| < 32$ mV, the peaks which avoid extinction ($\mathbf{q}_1$* and $\mathbf{q}_5$*) always become non-dispersive above Δ_0 (black arrow). An interesting precursor effect can be seen in the departure of the dispersive $\mathbf{q}_5(E)$ from expectations for $E < \Delta_0$. **f**, The decay of $Z(\mathbf{q}, E)$ intensity in all the dispersive modulation wavevectors and its disappearance below the noise level at the extinction energy Δ_0 (black arrow). These effects are described for all data sets in the Supplementary Information.

Outline of methods

Here we report the first application of these techniques to the simultaneous study of $\mathbf{r}$-space and $\mathbf{k}$-space electronic structure as $p \to 0$ in a copper oxide superconductor. We measure $g(\mathbf{r}, V)$ in ~50 nm × 50 nm fields of view in five $Bi_2Sr_2CaCu_2O_{8+\delta}$ samples, respectively with $p \approx$ 19%, 17%, 14%, 8%, 6% and transition temperatures T_c = 86 K, 88 K, 74 K, 45 K, 20 K. Each sample is inserted into the cryogenic ultrahigh vacuum of the scanning tunnelling microscope and cleaved to reveal an atomically clean BiO surface, and all $g(\mathbf{r}, V)$ measurements are made with atomic resolution and register at or below 4.2 K. This data set consists of $>10^6$ tunnelling spectra and is presented and analysed in detail in the Supplementary Information. From it, the $Z(\mathbf{r}, E)$ and $Z(\mathbf{q}, E)$ are calculated. For all energies at which dispersive $Z(\mathbf{r}, E)$ modulations exist, the octet model is used to find $\mathbf{k}_B(E)$ by means of equations (1)–(7). Because we require comprehensive internal consistency of all $\mathbf{q}_i(E)$ within the octet model, the resulting $\mathbf{k}_B(E)$ is heavily overdetermined (see Supplementary Information, section III, and Supplementary Fig. 2). For QPI analysis at p = 19%, we use the $g(\mathbf{r}, V)$ and $g(\mathbf{q}, V)$ data directly, both because tip elevation errors are minimal and for comparison with previous studies[16].

Extinction of BQP interference

In Fig. 2c we show a representative low-energy $Z(\mathbf{q}, E)$ (from a $p \sim 8\%$ sample). It reveals the expected 16 pairs of $\mathbf{q}$ vectors, which are consistent with each other within the octet model. The measured dispersion of each independent $\mathbf{q}$ vector is shown by open symbols in Fig. 2e. Equivalent data for all hole densities studied are shown in Supplementary Fig. 2. We find that the dispersion of octet model $\mathbf{q}$ vectors always stops, and that the intensity of all but two $\mathbf{q}$ vectors diminishes to zero, at some weakly doping-dependent excitation energy which we label $E = \Delta_0$. This is indicated by black arrows in Fig. 2e, f. In Fig. 2d we show a representative $Z(\mathbf{q}, E)$ for $E > \Delta_0$ (from the same $p \sim 8\%$ sample). Here we detect only two non-dispersive $\mathbf{q}$ vectors, labelled $\mathbf{q}_1$* and $\mathbf{q}_5$* in Fig. 2d, e. The equivalent pair of non-dispersive $\mathbf{q}$ vectors has been detected using ARPES in optimally doped $Bi_2Sr_2CaCu_2O_{8+\delta}$ (ref. 37) and in underdoped $Ca_{2-x}Na_xCuO_2Cl_2$ (ref. 31). Here, however, we find that this phenomenon occurs in all samples with $6\% < p < 19\%$ (the $\mathbf{q}_1$* and $\mathbf{q}_5$* data for all $E > \Delta_0$ are shown for these p values in Supplementary

Fig. 2). Our first key finding is then that, above some energy Δ_0, the dispersing BQP interference wavevectors $\mathbf{q}_i(E)$ always disappear, to be replaced by a completely non-dispersive excitation spectrum represented by $\mathbf{q}_1$* and $\mathbf{q}_5$*.

In Fig. 3a we plot the $\mathbf{k}_B(E)$ determined from the same data. Here it is apparent that when the BQP interference patterns disappear at Δ_0, the $\mathbf{k}$ states are near the diagonal lines between $(0, \pm\pi/a_0)$ and $(\pm\pi/a_0, 0)$ within the CuO_2 Brillouin zone. This $\mathbf{k}$-space 'extinction point' for BQP interference is defined not only by the change from dispersive to non-dispersive characteristics, but also by the disappearance of the $\mathbf{q}_2$, $\mathbf{q}_3$, $\mathbf{q}_6$ and $\mathbf{q}_7$ modulations (Fig. 2f and Supplementary Fig. 5). Thus, the BQP interference signatures of delocalized Cooper pairs vanish close to the perimeter of a $\mathbf{k}$-space region bounded by lines between $(0, \pm\pi/a_0)$ and $(\pm\pi/a_0, 0)$. We emphasize that this occurs neither at precisely the same $\mathbf{k}$ vector nor energy Δ_0 for each p, but always near the boundary of this restricted region (Fig. 3a). Within this region, the quasi-particles are confined to a Bogoliubov arc (fine solid lines in Fig. 3a) that shrinks rapidly towards $\mathbf{k} = (\pm 1/2, \pm 1/2)\pi/a_0$ with decreasing p. We hypothesize that this Bogoliubov arc is always coincident with the Fermi arc detected in the normal state[30–34,36,37].

Carrier density counts

Conventional theory would predict that the minima (maxima) of the Bogoliubov bands $\mathbf{k}_B(\pm E)$ should occur at the $\mathbf{k}$-space location of the Fermi surface of the non-superconducting state. By making this assumption here, we may ask if the carrier density count satisfies Luttinger's theorem, which states that twice the $\mathbf{k}$-space area enclosed by the Fermi surface, measured in units of the area of the first Brillouin zone, equals the number of electrons per unit cell, n. In Supplementary Fig. 4a we show as fine solid lines hole-like Fermi surfaces fitted to our measured $\mathbf{k}_B(E)$. Using Luttinger's theorem with these $\mathbf{k}$-space contours would result in a calculated hole density p (when measured from half filling, this is defined conventionally as $1-n$) for comparison with the estimated hole density in the samples. These data are shown by filled symbols in the inset to Fig. 3a. We see that the classic Luttinger theorem is strongly violated at all dopings below $p \sim 10\%$. This is neither a unique nor an anomalous observation: equivalent results have been reported previously for $Ca_{2-x}Na_xCuO_2Cl_2$ at similarly low p values[31].

However, the copper oxides are not metals but carrier-doped Mott insulators. For such systems, Luttinger's theorem must be amended[39] so that the zero-energy contours bounding the $\mathbf{k}$-space region representing carriers are defined not only by poles in the Green's functions, but also by their zeros. Essentially, the perturbation theory description of the metallic Fermi liquid breaks down and correlations among the particles generate zeros in the Green's functions. The locus of zeros of these Green's functions may be expected to occur at the lines joining $\mathbf{k} = (0, \pm\pi/a_0)$ to $\mathbf{k} = (\pm\pi/a_0, 0)$. In that situation, the hole density is related quantitatively to the area between the lines joining $\mathbf{k} = (0, \pm\pi/a_0)$ to $\mathbf{k} = (\pm\pi/a_0, 0)$ and the Fermi arcs. The carrier densities calculated using the region bounded by $\mathbf{k}_B(E)$ (arcs in Fig. 3a) and the hypothesized lines of zeros between $\mathbf{k} = (0, \pm\pi/a_0)$ and $\mathbf{k} = (\pm\pi/a_0, 0)$ (for example the dashed diagonal line in Fig. 3a) are shown by open symbols in the inset to Fig. 3a. These are in better agreement with the estimated hole density (see Supplementary Information). Thus, we conclude that if the Green's function lines of zeros occur between $\mathbf{k} = (0, \pm\pi/a_0)$ and $\mathbf{k} = (\pm\pi/a_0, 0)$, the measured $\mathbf{k}$-space structure and the doped-hole density can remain consistent as $p \rightarrow 0$.

Relationship of superconducting energy gap to the pseudogap energy

In Fig. 3b we plot the doping dependence of the superconducting energy gap $\Delta(\theta_\mathbf{k})$ (see Supplementary Information, section IV) in terms of $\theta_\mathbf{k}$, the angle in $\mathbf{k}$-space measured about the point (π, π) as shown in the inset. We find that $\Delta(\theta_\mathbf{k})$ is always cut off by the BQP extinction at the boundary of the restricted region. Moreover, we find

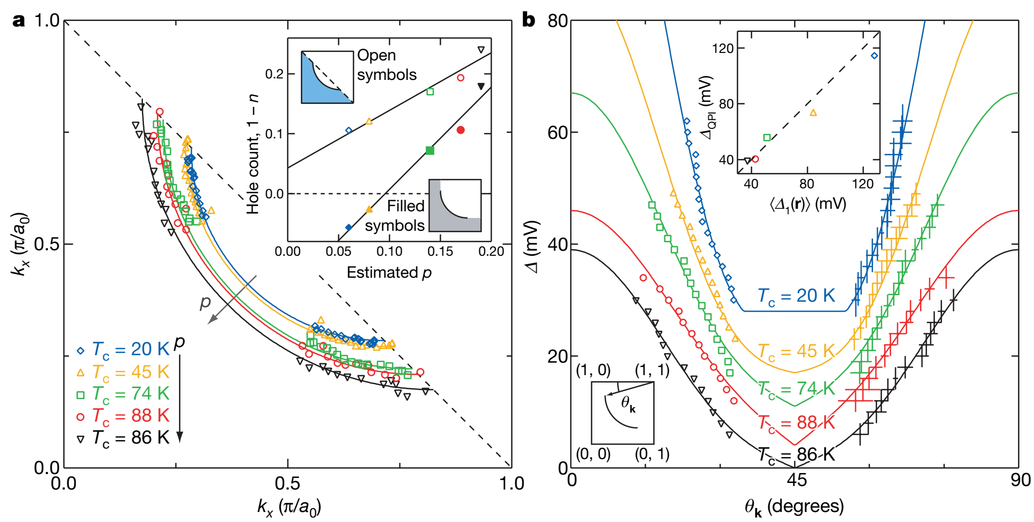


Figure 3 | Extinction of BQP interference. a, Locus of the Bogoliubov band minimum $\mathbf{k}_B(E)$ found from extracted QPI peak locations $\mathbf{q}_i(E)$, in five independent $Bi_2Sr_2CaCu_2O_{8+\delta}$ samples with decreasing hole density. Fits to quarter-circles are shown and, as p decreases, these curves enclose a progressively smaller area. We find that the BQP interference patterns disappear near the perimeter of a $\mathbf{k}$-space region bounded by the lines joining $\mathbf{k} = (0, \pm\pi/a_0)$ and $\mathbf{k} = (\pm\pi/a_0, 0)$. The spectral weights of $\mathbf{q}_2$, $\mathbf{q}_3$, $\mathbf{q}_6$ and $\mathbf{q}_7$ vanish at the same place (dashed line; see also Supplementary Fig. 3). Filled symbols in the inset represent the hole count $p = 1 - n$ derived using the simple Luttinger theorem, with the fits to a large, hole-like Fermi surface shown in Supplementary Fig. 4a and indicated schematically here in grey. Open symbols in the inset are the hole counts calculated using the area enclosed by the Bogoliubov arc and the lines joining $\mathbf{k} = (0, \pm\pi/a_0)$ and $\mathbf{k} = (\pm\pi/a_0, 0)$, and are indicated schematically here in blue. **b**, The evolution of the superconducting energy gap $\Delta(\mathbf{k})$ is shown for the $\mathbf{k}$-space points shown in **a**, but here as a function of the angle $\theta_\mathbf{k}$ about the point (π, π) (see the lower insert). These are fitted to the parameterization $\Delta(\theta_\mathbf{k}) = \Delta_{QPI}[B\cos(2\theta_\mathbf{k}) + (1 - B)\cos(6\theta_\mathbf{k})]$ for each hole density (and offset vertically for clarity) as indicated by the fine solid lines. The measured $\Delta(\theta_\mathbf{k})$ values are shown by the coloured data symbols on the left-hand side, and the error bars (1 s.d.) for each measurement are shown on the right-hand side. As p decreases, it is obvious that the fitted Δ_{QPI} increases rapidly. The upper inset shows the relationship between Δ_{QPI} and the average pseudogap energy $\langle\Delta_1\rangle$.

that the energy $\Delta_0(p)$ where BQP interference disappears and the spatially averaged energy $\bar{\Delta}_0(p)$ at which electronic homogeneity is lost in $Bi_2Sr_2CaCu_2O_{8+\delta}$ (ref. 26) are indistinguishable within their uncertainties. The function $\Delta(\theta_{\mathbf{k}}) = \Delta_{QPI}[B\cos(2\theta_{\mathbf{k}}) + (1 - B)\cos(6\theta_{\mathbf{k}})]$, where Δ_{QPI} is the theoretical superconducting energy gap maximum at $\theta_{\mathbf{k}} = 0, \pi/2$, is required to fit the measured $\Delta(\theta_{\mathbf{k}})$, as shown by the fine solid lines in Fig. 3b. We find that as p decreases, Δ_{QPI} increases rapidly and B decreases slowly. Finally, the maximum energy of the fitted superconducting gap Δ_{QPI} is always in good quantitative agreement with the spatially averaged pseudogap maxima $\langle\Delta_1\rangle$ as derived from the particle–hole-symmetric peaks in the spectra (Fig. 1b); this relationship is shown in the inset to Fig. 3b.

r-space structure of pseudogap excitations

Next we examine the structure of excitations above the extinction energy Δ_0, where no dispersive QPI is detected. We find that these $Z(\mathbf{q}, E)$ have only two non-dispersive $\mathbf{q}$ vectors, namely $\mathbf{q}_1$* and $\mathbf{q}_5$*, which evolve with p as shown implicitly in Fig. 3a (and in detail in Supplementary Fig. 6). As might be expected from their lack of energy dispersion, it is in **r**-space that these excitations appear most well defined. Analysis of $Z(\mathbf{r}, E)$ for $\Delta_0 < E < 150$ meV shows spatial patterns that are highly similar at all energies but have spatial variations in intensity. Representative examples are shown in Fig. 4a, b. The patterns are short-correlation-length Cu–O–Cu bond-centred modulations in $Z(\mathbf{r}, E)$ with nanoscale unidirectional domains $\sim 4a_0$ wide embedded in a glassy matrix. The spatial structure in these **r**-space patterns (Fig. 4a, b) appears closely related to that detected by maps of:

$$R(\mathbf{r}, E = eV) \equiv \frac{I(\mathbf{r}, +V)}{I(\mathbf{r}, -V)}$$

These quantify variations in the energy-integrated tunnelling asymmetry, as described in ref. 20 for $V = 150$ meV; their spatial arrangement forms a Cu–O–Cu bond-centred electronic pattern with dispersed $\sim 4a_0$-wide unidirectional nano-domains. However, because these maps integrate over energy, they do not reveal the characteristic energy of the constituent **r**-space phenomena.

To address this issue, we focus on the maximum intensity of $Z(\mathbf{r}, E)$ for each E. This fluctuates strongly in space as shown, for example, in Fig. 4a, b. However, simultaneous images of the pseudogap energy scale $\Delta_1(\mathbf{r})$ (as defined in the inset to Fig. 4d) also show strong spatial fluctuations (Fig. 4c). Comparing these with Fig. 4a, b, it seems that $Z(\mathbf{r}, E)$ exhibits its maximum intensity in the spatial regions where $E = \Delta_1(\mathbf{r})$. To quantify this, we scale the energy E at each **r** by the pseudogap magnitude $\Delta_1(\mathbf{r})$ at the same location, thus defining the new energy scale $e(\mathbf{r}) = E/\Delta_1(\mathbf{r})$ to be a fraction of the local pseudogap energy scale. We find that the translational- and C_4-symmetry-breaking bond-centred modulations exhibit an apparent maximum intensity at $e = 1$, or $E(\mathbf{r}) = \Delta_1(\mathbf{r})$ (Supplementary Information, section VII, and Supplementary Fig. 7). Our conclusion is then that the intricate **r**-space patterning of electronic structure seen in the maps of R (ref. 20 and Fig. 4f) is actually an atomic-scale visualization of the spatial structure of low-p pseudogap excitations (Figs 4e and 5a and Supplementary Fig. 7).

Summary and discussion

As p is reduced towards the Mott insulator state, scattering interference modulations of BQPs always disappear at an energy Δ_0 that is indistinguishable from the energy at which electronic homogeneity is lost[26]. BQP interference disappears near the perimeter of a region in **k**-space restricted by the lines joining $\mathbf{k} = (0, \pm\pi/a_0)$ and $\mathbf{k} = (\pm\pi/a_0, 0)$. For energies $E > \Delta_0$, the electronic structure appears to be static in **r**-space and independent of E. In fact, it consists of the atomic-scale spatial patterns previously reported[20] but here identified as the pseudogap excitations at $E = \pm\Delta_1$. Our observations therefore provide a new and different context within which to understand the two excitation energy scales as $p \to 0$. The lower energy, Δ_0, is associated with the disappearance of the BQP interference arising from the presence of delocalized Cooper pairs, whereas the upper energy, Δ_1, is associated with the characteristic **r**-space electronic structure of the pseudogap excitations. Overall, a progressive conversion from the former to the latter electronic structure occurs as p decreases to zero even though their characteristic energies Δ_{QPI} and Δ_1 remain equal. Perhaps most notably, the low-p pseudogap excitations locally break the translational symmetry, and reduce the C_4 symmetry of the electronic structure in each four-Cu-atom plaquette to C_2 symmetry in Cu–O–Cu bond-centred patterns without long-range order[20].

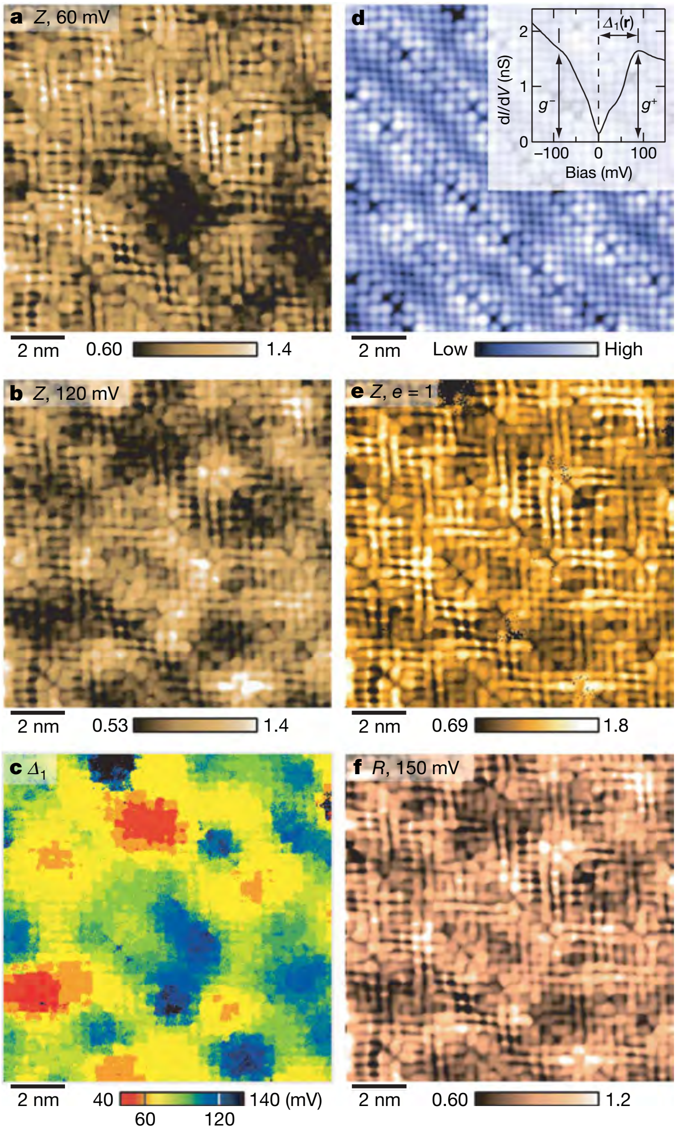


Figure 4 | Imaging copper oxide pseudogap excitations as $p \to 0$. **a**, **b**, Atomically resolved $Z(\mathbf{r}, E)$ for the $T_c = 45$ K sample (simultaneous topographic image shown in **d**) for two distinct energies (60, 120 mV). The intensity of electronic structure patterns in **r**-space vary as a function of energy. **c**, The corresponding spatial map of the Δ_1 gap magnitude over the sample surface in **d**. There is a wide distribution of heterogeneous pseudogap values[26]. **d**, A topograph showing the locations of Bi atoms (small, bright circles) on the field of view where all data in this figure were acquired. The inset shows the local definition of $\Delta_1(\mathbf{r})$ with the tunnelling conductances at $E = \pm\Delta_1$ represented by g^+ and g^-, respectively. **e**, Image of $Z(\mathbf{r}, e = E/\Delta_1 = 1)$ in which energy has been rescaled by the local value of Δ_1(r) from **c**; this represents an image of what pseudogap states would look like in terms of $Z(\mathbf{r}, E = \Delta_1)$ if the nanoscale disorder in Δ_1 were not to exist. **f**, The $R(\mathbf{r}, E = 150$ mV$)$ patterns are virtually identical to those in **e**. Thus, the spatial patterning reported in ref. 20 is actually concentrated on the states at $E = \pm\Delta_1$, meaning that these **r**-space excited states are the copper oxide pseudogap excitations as $p \to 0$.

A number of theoretical issues emerge from these observations. For example, we do not know why the BQP interference extinction occurs near the perimeter of the **k**-space region bounded by the lines joining $\mathbf{k} = (0, \pm\pi/a_0)$ and $\mathbf{k} = (\pm\pi/a_0, 0)$. One reason could be that this **k**-space perimeter coincides with the boundary at which elastic particle–particle Umklapp scattering should intensify rapidly as the Mott insulator state is approached[1]. A different possibility is that the **r**-space electronic structure has undergone a $\sqrt{2} \times \sqrt{2}$ reconstruction in which the periodicity of the crystal structure is increased owing to the appearance of a coexisting long-range-ordered state; the arcs in Fig. 3a would then represent a hole pocket within a reduced Brillouin zone. However, neither the antiferromagnetism nor the other electronic long-range orders[40,41] necessary for such a reconstruction are observed in $Bi_2Sr_2CaCu_2O_{8+\delta}$. A related explanation could be inelastic scattering of the quasi-particles by spin fluctuations[42,43] at $\mathbf{Q} = (\pi, \pi)$ or by fluctuations of a circulating-current order[40,41] (each would exhibit a $\sqrt{2} \times \sqrt{2}$ reconstruction if stabilized). A key difficulty with all these possibilities is the loss of translational symmetry and local reduction of the expected C_4 symmetry of the Cu plaquette to C_2 (Fig. 4a, b, e, f and ref. 20), which are not required within such models. Yet another explanation could be static spin–charge stripe glass coexisting with superconductivity[44–47]. This could explain the loss of translational symmetry and the reduction of C_4 symmetry to C_2 within the pseudogap excitations. However, it does not explain the locus of quasi-particle extinction along the lines joining $\mathbf{k} = (0, \pm\pi/a_0)$ and $\mathbf{k} = (\pm\pi/a_0, 0)$. A related class of proposals, perhaps more congruent with our data, suggest that the fluctuations of an order parameter that would break translational and C_4 symmetry if stabilized scatter the quasi-particles strongly[26] and alter the QPI processes[48–50] in a way consistent with our observations.

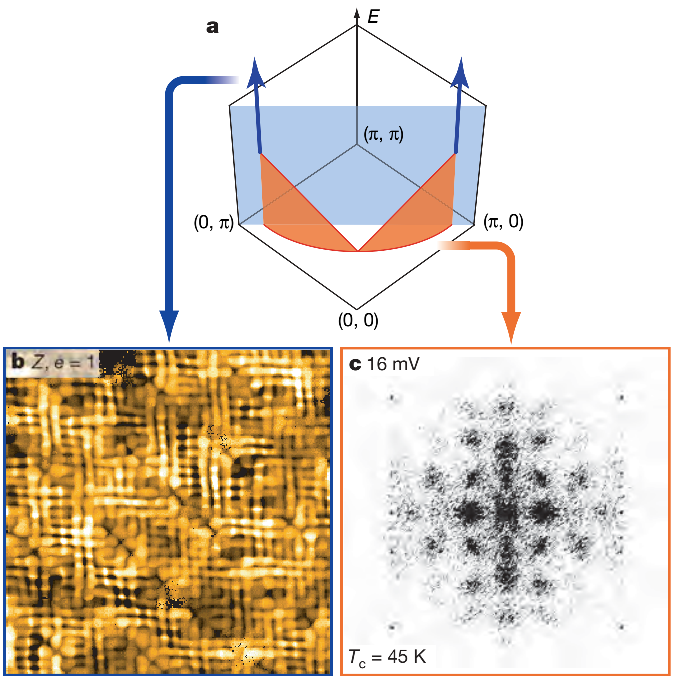


Figure 5 | Bipartite electronic structure of copper oxides as $p \to 0$. **a**, The pseudogap excitations are separated from the Bogoliubov quasi-particle states by a plane in E–**k** space restricted approximately by the lines joining $\mathbf{k} = (0, \pm\pi/a_0)$ and $\mathbf{k} = (\pm\pi/a_0, 0)$. This is indicated in blue. The energy $E = \Delta_0$ at which the excitations make the transition from being dispersive BQP to being energy-independent pseudogap states (and at which spatial homogeneity is lost[26]), diminishes slowly with decreasing p. **b**, For energies $E > \Delta_0$, the electronic structure is static in **r**-space and independent of E; the pseudogap states at $|E| = \Delta_1$ are short-correlation-length Cu–O–Cu bond-centred modulations in $Z(\mathbf{r}, E)$ that break translational symmetry and reduce C_4 symmetry to C_2 within each four-Cu-atom plaquette (see ref. 20). **c**, In contrast, scattering interference between **k**-space BQP excitations from delocalized Cooper pairs is observed on an arc terminated approximately by the plane in E–**k** space (blue in **a**) that connects $\mathbf{k} = (0, \pi/a_0)$ to $\mathbf{k} = (\pi/a_0, 0)$. As $p \to 0$, the arc shrinks and the phenomena in **c** rapidly disappear, to be replaced by those in **b**.

Notwithstanding these theoretical issues, a new empirical model of the bipartite electronic structure of copper oxides emerges here (Fig. 5). Its key elements include the following: as p decreases, the BQP interferences indicating the presence of delocalized Cooper pairs disappear near the lines connecting $(\pm\pi, 0)$ and $(0, \pm\pi)$ in **k**-space, even as the Cooper pairing energy continues to rise; although the pseudogap excitations appear to have the same energy scale, Δ_1, as the maximum energy scale of superconductivity, Δ_{QPI}, they have a radically different **r**-space phenomenology that locally breaks the C_4 symmetry on Cu plaquettes down to C_2 in a bond-centred pattern. There are two immediate tests of the validity of this new model. First, C_4 symmetry should be recovered on each Cu plaquette at even lower hole density as the antiferromagnetism reappears. Second, the diverse repercussions of this bipartite electronic structure for the bulk thermodynamic characteristics should be compared with the results of relevant measurements.

Received 7 April; accepted 3 July 2008.

1. Honerkamp, C., Salmhofer, M., Furukawa, N. & Rice, T. M. Breakdown of the Landau-Fermi liquid in two dimensions due to Umklapp scattering. *Phys. Rev. B* **63,** 035109 (2001).
2. Paramekanti, A., Randeria, M. & Trivedi, N. Projected wave functions and high temperature superconductivity. *Phys. Rev. Lett.* **87,** 217002 (2001).
3. Altman, E. & Auerbach, A. Plaquette boson-fermion model of cuprates. *Phys. Rev. B* **65,** 104508 (2002).
4. Sénéchal, D. & Tremblay, A.-M. S. Hot spots and pseudogaps for hole- and electron-doped high-temperature superconductors. *Phys. Rev. Lett.* **92,** 126401 (2004).
5. Civelli, M., Capone, M., Kancharla, S. S., Parcollet, O. & Kotliar, G. Dynamical breakup of the Fermi surface in a doped Mott insulator. *Phys. Rev. Lett.* **95,** 106402 (2005).
6. Stanescu, T. D. & Kotliar, G. Fermi arcs and hidden zeros of the Green function in the pseudogap state. *Phys. Rev. B* **74,** 125110 (2006).
7. Berthod, C., Giamarchi, T., Biermann, S. & Georges, A. Breakup of the Fermi surface near the Mott transition in low-dimensional systems. *Phys. Rev. Lett.* **97,** 136401 (2006).
8. Civelli, M. *et al.* Nodal-antinodal dichotomy and the two gaps of a superconducting doped Mott insulator. *Phys. Rev. Lett.* **100,** 046402 (2008).
9. Haule, K. & Kotliar, G. Strongly correlated superconductivity: A plaquette dynamical mean-field theory study. *Phys. Rev. B* **76,** 104509 (2007).
10. Konik, R. M., Rice, T. M. & Tsvelik, A. M. Doped spin liquid: Luttinger sum rule and low temperature order. *Phys. Rev. Lett.* **96,** 086407 (2006).
11. Yang, K.-Y., Rice, T. M. & Zhang, F.-C. Phenomenological theory of the pseudogap state. *Phys. Rev. B* **73,** 174501 (2006).
12. Stanescu, T. D., Phillips, P. & Choy, T.-P. Theory of the Luttinger surface in doped Mott insulators. *Phys. Rev. B* **75,** 104503 (2007).
13. Norman, M. R., Kanigel, A., Randeria, M., Chatterjee, U. & Campuzano, J. C. Modeling the Fermi arc in underdoped cuprates. *Phys. Rev. B* **76,** 174501 (2007).
14. Crommie, M. F., Lutz, C. P. & Eigler, D. M. Imaging standing waves in a two-dimensional electron gas. *Nature* **363,** 524–527 (1993).
15. Wang, Q.-H. & Lee, D.-H. Quasiparticle scattering interference in high-temperature superconductors. *Phys. Rev. B* **67,** 020511 (2003).
16. McElroy, K. *et al.* Relating atomic-scale electronic phenomena to wave-like quasiparticle states in superconducting $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Nature* **422,** 592–596 (2003).
17. Capriotti, L., Scalapino, D. J. & Sedgewick, R. D. Wave-vector power spectrum of the local tunneling density of states: Ripples in a *d*-wave sea. *Phys. Rev. B* **68,** 014508 (2003).
18. Nunner, T. S., Chen, W., Andersen, B. M., Melikyan, A. & Hirschfeld, P. J. Fourier transform spectroscopy of *d*-wave quasiparticles in the presence of atomic scale pairing disorder. *Phys. Rev. B* **73,** 104511 (2006).
19. Hanaguri, T. *et al.* Quasiparticle interference and superconducting gap in $Ca_{2-x}Na_xCuO_2Cl_2$. *Nature Phys.* **3,** 865–871 (2007).
20. Kohsaka, Y. *et al.* An intrinsic bond-centered electronic glass with unidirectional domains in underdoped cuprates. *Science* **315,** 1380–1385 (2007).
21. Orenstein, J. & Millis, A. J. Advances in the physics of high-temperature superconductivity. *Science* **288,** 468–474 (2000).
22. Hüefner, S., Hossain, M. A., Damascelli, A. & Sawatzky, G. A. Two gaps make a high-temperature superconductor? *Rep. Prog. Phys.* **71,** 062501 (2008).
23. Fischer, Ø., Kugler, M., Maggio-Aprile, I., Berthod, C. & Renner, C. Scanning tunneling spectroscopy of high-temperature superconductors. *Rev. Mod. Phys.* **79,** 353–419 (2007).

24. Lang, K. M. *et al.* Imaging the granular structure of high-T_c superconductivity in underdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Nature* **415,** 412–416 (2002).
25. McElroy, K. *et al.* Coincidence of checkerboard charge order and antinodal state decoherence in strongly underdoped superconducting $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Phys. Rev. Lett.* **94,** 197005 (2005).
26. Alldredge, J. W. *et al.* Evolution of the electronic excitation spectrum with strongly diminishing hole-density in superconducting $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Nature Phys.* **4,** 319–326 (2008).
27. Deutscher, G. Coherence and single-particle excitations in the high-temperature superconductors. *Nature* **397,** 410–412 (1999).
28. Gedik, N., Orenstein, J., Liang, R., Bonn, D. A. & Hardy, W. N. Diffusion of nonequilibrium quasi-particles in a cuprate superconductor. *Science* **300,** 1410–1412 (2003).
29. Tacon, M. L. *et al.* Two energy scales and two distinct quasiparticle dynamics in the superconducting state of underdoped cuprates. *Nature Phys.* **2,** 537–543 (2006).
30. Norman, M. R. *et al.* Destruction of the Fermi surface in underdoped high-T_c superconductors. *Nature* **392,** 157–160 (1998).
31. Shen, K. M. *et al.* Nodal quasiparticles and antinodal charge ordering in $Ca_{2-x}Na_xCuO_2Cl_2$. *Science* **307,** 901–904 (2005).
32. Tanaka, K. *et al.* Distinct Fermi-momentum-dependent energy gaps in deeply underdoped Bi2212. *Science* **314,** 1910–1913 (2006).
33. Kanigel, A. *et al.* Evolution of the pseudogap from Fermi arcs to the nodal liquid. *Nature Phys.* **2,** 447–451 (2006).
34. Kanigel, A. *et al.* Protected nodes and the collapse of Fermi arcs in high-T_c cuprate superconductors. *Phys. Rev. Lett.* **99,** 157001 (2007).
35. Anderson, P. W. *et al.* The physics behind high-temperature superconducting cuprates: the 'plain vanilla' version of RVB. *J. Phys. Condens. Matter* **16,** R755–R769 (2004).
36. Chatterjee, U. *et al.* Nondispersive Fermi Arcs and the absence of charge ordering in the pseudogap phase of $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Phys. Rev. Lett.* **96,** 107006 (2006).
37. Chatterjee, U. *et al.* Anomalous dispersion in the autocorrelation of angle-resolved photoemission spectra of high-temperature $Bi_2Sr_2CaCu_2O_{8+\delta}$ superconductors. *Phys. Rev. B* **76,** 012504 (2007).
38. McElroy, K. *et al.* Elastic scattering susceptibility of the high temperature superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$: A comparison between real and momentum space photoemission spectroscopies. *Phys. Rev. Lett.* **96,** 067005 (2006).
39. Dzyaloshinskii, I. Some consequences of the Luttinger theorem: The Luttinger surfaces in non-Fermi liquids and Mott insulators. *Phys. Rev. B* **68,** 085113 (2003).
40. Lee, P. A., Nagaosa, N. & Wen, X.-G. Doping a Mott insulator: Physics of high-temperature superconductivity. *Rev. Mod. Phys.* **78,** 17–85 (2006).
41. Chakravarty, S., Laughlin, R. B., Morr, D. K. & Nayak, C. Hidden order in the cuprates. *Phys. Rev. B* **63,** 094503 (2001).
42. Lee, W.-C., Sinova, J., Burkov, A. A., Joglekar, Y. & MacDonald, A. H. Theory of reduced superfluid density in underdoped cuprate superconductors. *Phys. Rev. B* **77,** 214518 (2008).
43. Graser, S., Hirschfeld, P. J. & Scalapino, D. J. Local quasiparticle lifetimes in a *d*-wave superconductor. *Phys. Rev. B* **77,** 184504 (2008).
44. Balents, L., Bartosch, L., Burkov, A., Sachdev, S. & Sengupta, K. Putting competing orders in their place near the Mott transition. II. The doped quantum dimer model. *Phys. Rev. B* **71,** 144509 (2005).
45. Tešanović, Z. *d*-wave duality and its reflections in high-temperature superconductors. *Nature Phys.* **4,** 408–414 (2008).
46. Vojta, M. & Rösch, O. Superconducting *d*-wave stripes in cuprates: Valence bond order coexisting with nodal quasiparticles. *Phys. Rev. B* **77,** 094504 (2008).
47. Granath, M. Fermi momentum resolved charge order for spin-disordered stripes. *Phys. Rev. B* **77,** 165128 (2008).
48. Kim, E.-A. *et al.* Theory of the nodal nematic quantum phase transition in superconductors. *Phys. Rev. B* **77,** 184514 (2008).
49. Pelissetto, A., Sachdev, S. & Vicari, E. Nodal quasiparticles and the onset of spin density wave order in the cuprates. *Phys. Rev. Lett.* **101,** 027005 (2008).
50. Vojta, M. Electronic properties of disordered valence-bond stripes in cuprate superconductors. Preprint at ⟨http://arxiv.org/abs/0803.2038⟩ (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge and thank A. V. Balatsky, J. C. Campuzano, E. Fradkin, A. Georges, T. Hanaguri, P. J. Hirschfeld, S. Kivelson, E.-A. Kim, G. Kotliar, P. A. Lee, M. Norman, P. Phillips, M. Randeria, T. M. Rice, S. Sachdev, K. Shen, Z. X. Shen, A. Tsvelik, M. Vojta and F. C. Zhang for discussions. This work was supported by the US National Science Foundation through the Cornell Center for Material Research, by Brookhaven National Laboratory, by the US Department of Energy, by the US Office of Naval Research, by a Grant-in-Aid for Scientific Research from the Ministry of Science and Education (Japan), and by the 21st Century COE Program of the Japan Society for the Promotion of Science. P.W. acknowledges support from the Humboldt Foundation and A.S. acknowledges support from the US Army Research Office.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.C.D. (jcdavis@ccmr.cornell.edu).

Vol 454 | 28 August 2008 | doi:10.1038/nature07155

Egalitarianism in young children

Ernst Fehr[1,2], Helen Bernhard[1] & Bettina Rockenbach[3]

Human social interaction is strongly shaped by other-regarding preferences, that is, a concern for the welfare of others. These preferences are important for a unique aspect of human sociality—large scale cooperation with genetic strangers—but little is known about their developmental roots. Here we show that young children's other-regarding preferences assume a particular form, inequality aversion that develops strongly between the ages of 3 and 8. At age 3–4, the overwhelming majority of children behave selfishly, whereas most children at age 7–8 prefer resource allocations that remove advantageous or disadvantageous inequality. Moreover, inequality aversion is strongly shaped by parochialism, a preference for favouring the members of one's own social group. These results indicate that human egalitarianism and parochialism have deep developmental roots, and the simultaneous emergence of altruistic sharing and parochialism during childhood is intriguing in view of recent evolutionary theories which predict that the same evolutionary process jointly drives both human altruism and parochialism.

Other-regarding preferences are decisive for the human ability to achieve and maintain cooperation in large groups of genetic strangers[1,2]. If an individual cares for the welfare of other group members, he or she is more likely to refrain from free-riding in cooperative projects. Similarly, if an individual dislikes the free-riding of others—because it is associated with inequality[3–5] or because it represents a norm violation[6]—the individual is more likely to punish free-riders[7–9]. This punishment then constitutes an incentive for potential free-riders to cooperate. Other-regarding preferences are also important in public life and politics[10] and they powerfully amplify reputational incentives in strategic interactions, thus contributing to the cooperation-enhancing force of reputation opportunities[11–14].

The developmental origins and proximate mechanisms behind other-regarding preferences are not well understood, however, despite recent progress[15–22]. Because we know little about when young children start to take the welfare of others into account, we conducted experiments with 229 young, genetically unrelated, Swiss children (127 girls and 102 boys) between the ages of 3 and 8. An understanding of the development of other-regarding preferences in children may enable us to gain deeper insights into the proximate and ultimate sources of species differences in preferences and cooperation. The study of children's preferences is also of particular interest in light of recent experiments in non-human primates[23–27], allowing a more direct comparison between human and non-human primates. Experiments with non-human primates have, for example, demonstrated that chimpanzees show little willingness to provide food to a familiar conspecific in situations where they could do so with no or only a small cost[23,24,27]. In view of this result, it is interesting to study whether (and if so, when) children are willing to provide valuable resources to their partners. In this way, the large species differences in cooperation between humans and non-human primates can be more directly traced back to species differences in other-regarding preferences.

Investigation of the development of moral judgment[28] and prosocial behaviour[29–35] is a rich tradition in psychology, but there is a surprising lack of studies that isolate the development of other-regarding preferences from the development of other forms of prosocial behaviour. The experimental study of other-regarding preferences in humans involves the conduct of one-shot experiments with anonymous interaction partners because the behaviour in non-anonymous face-to-face interactions, or in repeated interactions with the same partner, can easily be affected by selfish motives. A subject could, for example, behave prosocially because of the expectation of future benefits from the partner that accrue as a result of prosocial behaviour in the current interaction. Selfish motives could therefore drive prosocial behaviours such as sharing a valuable resource in non-anonymous face-to-face interactions or in repeated interactions between the experimental subjects. Measuring other-regarding preferences without such confounds thus requires the conduct of anonymous one-shot experiments.

Testing for inequality aversion in children

For this reason, we conducted experiments with young children that enabled us to measure other-regarding preferences such as inequality aversion. In the context of our experiments, inequality aversion prevails if subjects prefer allocations that reduce the inequality between themselves and their partner, regardless of whether the inequality is to their advantage or disadvantage[3].

Each subject participated in the three treatments described below and was paired with one other anonymous partner in each treatment. Each treatment condition was explained in detail to the decision maker so that we could be sure that the child had completely understood the experiment and the consequences of the different choices (see Methods and Supplementary Methods). In all treatments, the decision maker allocated units of sweets (smarties, jellybabies or fizzers) to himself and/or to the partner. In the 'prosocial' treatment, which was inspired by recent experiments with chimpanzees[23,24], the subject could choose between the allocation (1,1)—that is, 1 for himself, 1 for partner—and the allocation (1,0). This treatment measures some elementary form of prosociality, because by choosing (1,1) the subject can at no cost to himself deliver a benefit to the partner and, thus, avoid advantageous inequality. In principle, the choice of (1,1) can be driven by the equality motive[3] or by a motive to increase the partner's payoff or both parties joint payoff[36]. Economic self-interest is not involved in the prosocial game because the decision maker receives one unit regardless of which choice he makes. It is therefore also possible that a selfish individual who does not care about the partner's payoff will choose (1,1). In fact, because there is

[1]University of Zurich, Institute for Empirical Research in Economics, Blumlisalpstrasse 10, CH-8006 Zurich, Switzerland. [2]Collegium Helveticum, Schmelzbergstrasse 25, CH-8092 Zurich, Switzerland. [3]University of Erfurt, Nordhäuser Straße 63, D-99089 Erfurt, Germany.

no reason for a selfish individual to make either choice, a population of self-interested individuals would choose (1,1) in 50% of the cases. For this reason, evidence for other-regarding behaviour in the prosocial game requires that the population of children choose (1,1) significantly above 50% frequency.

In the 'envy' treatment, the subject could choose between (1,1) and (1,2). Here again, it is possible to deliver a benefit to the partner at no cost, but the choice (1,2) leads to disadvantageous inequality for the decision maker. Thus, if an individual just wants to increase the partner's or the joint payoff, he should choose (1,1) in the prosocial treatment and (1,2) in the envy treatment. In contrast, if the equality motive drives behaviour in these two conditions, the subject chooses (1,1) in both treatments, thus avoiding the unequal allocations (1,0) in the prosocial treatment and (1,2) in the envy treatment. However, as in the prosocial treatment, a purely selfish individual has no reason to make either choice in the envy treatment. For this reason, evidence for the equality motive in the envy game again requires that the population of children choose (1,1) significantly above 50% frequency.

In a third condition, the 'sharing' treatment, the subject could choose between (1,1) and (2,0). This treatment measures a strong form of inequality aversion because the provision of a benefit for the partner is costly for the subject. Selfish children should therefore never make the egalitarian choice in this treatment, indicating that the choice of (1,1) unambiguously suggests an other-regarding preference. The sharing treatment also enables us to measure altruism as defined by evolutionary biology because sharing implies a costly transfer of a valued resource to another individual.

In addition to these treatments, we also implemented an ingroup and an outgroup condition 'across subjects'. In the ingroup condition, the partner came from the same playschool, kindergarten or school, whereas the partner came from another playschool, kindergarten or school in the outgroup condition. The rationale for the outgroup condition is provided by evidence and theory indicating that parochialism strongly shapes adult human altruism[37–39], and that the same evolutionary process might determine the development of both human altruism and parochialism[40].

From self-interest to inequality aversion

Among the 3–4-yr-old children, most behaved selfishly in the ingroup condition of the sharing game because only 8.7% of the children were willing to share (Fig. 1). The fact that the frequency of (1,1) choices did not differ significantly from 50% in the prosocial game and in the envy game further supports the low incidence of other-regarding preferences at this age (binomial test; $P = 0.21$ for the prosocial game, $P = 0.68$ for the envy game; $n = 23$). The prevalence of selfish behaviour in the sharing game decreased slightly for 5–6-yr-old children, but 78% were still not willing to share at this age. Also, as with the 3–4-yr-old children, the frequency of egalitarian choices in the prosocial and the envy game for the 5–6-yr-olds was not significantly different from 50% (binomial test; $P = 0.24$ for the prosocial game, $P = 0.41$ for the envy game; $n = 36$). A substantially different picture emerged, however, for children at ages 7–8 (Fig. 1): 45% of them showed sharing behaviour, and we also found strong evidence for other-regarding preferences in the other two games. Of the 7–8-yr-olds, 78% preferred the egalitarian allocation in the prosocial game, refuting the null hypothesis of random choices (binomial test, $P < 0.001$, $n = 56$). Similarly, an overwhelming majority (80%) preferred the egalitarian alternative in the envy game at this age (binomial test, $P < 0.001$, $n = 56$).

Taken together, the behavioural patterns across all three games suggest that children at age 3–4 show little willingness to share resources but a non-negligible percentage of the children is willing to make choices that benefit the recipient if it is not costly. After this age, other-regarding preferences develop, which take the form of inequality aversion instead of a simple preference for increasing the partner's or the joint payoff. If the motive to increase the partner's or the joint payoff were to drive the children's other-regarding preferences, they would have then chosen the alternative (1,2) in the envy game. In fact, however, most of the children at age 7–8 preferred the egalitarian allocation.

Therefore, if we pool the children's choices across the various games, we find that both strongly and weakly egalitarian choices show a large increase with age (Fig. 2, Supplementary Fig. 3 and Supplementary Table 1). Egalitarianism, which is characterized by a (1,1) choice both in the prosocial and the envy game (red columns in Fig. 2), increases from 21% at age 3–4 to 33% at age 5–6, whereas 60% prefer the egalitarian allocation in both games at age 7–8. The percentage of egalitarian choices at age 7–8 significantly differs from an independent random choice in each of both games (binomial test, $P < 0.0001$, $n = 56$). If the children had made independent, random choices, only 25% of them would have chosen (1,1) in both the prosocial and the envy game. If we pool the children's choices in all three games, the percentage of children who preferred the egalitarian allocation in all three games increases from 4% at age 3–4 to 30% at age 7–8 (Fig. 2). Thus, among those children who choose the egalitarian allocation in the prosocial and the envy game at age 7–8, roughly 50% share resources in the sharing game. The other 50% choose the selfish allocation (2,0) in the sharing game.

It is notable that the share of subjects who maximize the partner's payoff by choosing both (1,1) in the prosocial game and (1,2) in the envy game (blue columns in Fig. 2) decreases sharply from 43% at age 3–4 to 16% at age 7–8. Furthermore, the percentage of subjects who maximize the partner's payoff in all three games is only roughly 5% and does not change much with age. The across-game perspective also enables us to identify a third type of subject whom we call 'spiteful' because they minimize the partner's payoff in all three games. The share of spiteful subjects is 22% at age 3–4 and 5–6 and decreases slightly to 14% at age 7–8, a percentage that is similar to the relative share of spiteful subjects observed in adult subject pools[41,42].

Parochial egalitarianism

Parochial tendencies affected children's choices in all three treatments, and these tendencies are pervasive in the sharing and prosocial

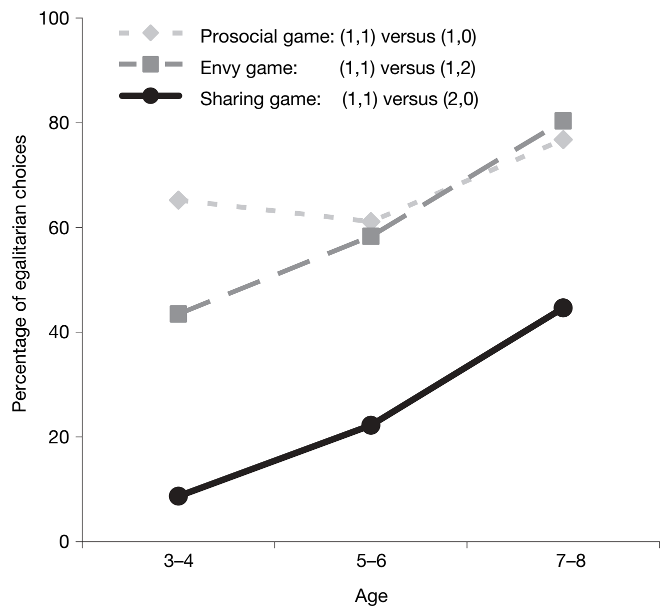


Figure 1 | The relative frequency of egalitarian choices across all ingroup treatments. In these treatments, the decision maker's choice determines the resources of an ingroup partner. The frequency of egalitarian choices strongly increases with age across all three ingroup treatments, and most children prefer equality at age 7–8 in the prosocial and the envy game. However, if equality is costly for the children, they choose the egalitarian allocation less frequently—as indicated by the behaviour in the sharing game—and at age 3–4, self-interested choices dominate almost completely.

games (Fig. 3). The egalitarian choice is roughly 15%–20% more likely in the prosocial game if the partner is an ingroup member (Fig. 3a). This difference is highly significant (ingroup dummy in probit regression, $P = 0.004$, $z = 2.92$, $n = 229$) and a similar ingroup–outgroup gap prevails across all ages. The largest difference was found in the sharing game (Fig. 3b), where we observed a strong increase in the frequency of egalitarian choices if the partner was from the ingroup (age effect in probit regression, $P = 0.001$, $z = 3.33$, $n = 115$), whereas the children's willingness to share even slightly declined with age in the outgroup condition although this was not significant (age effect in probit regression, $P = 0.123$, $z = -1.54$, $n = 114$). We did observe, however, a highly significant interaction effect between age and outgroup condition (probit regression, $P < 0.001$, $z = 3.50$, $n = 229$), indicating that the difference between sharing in the ingroup and outgroup conditions strongly increases with age. Thus, the children's altruism and parochialism emerges simultaneously between the ages of 3 and 8 and is associated with a very strong ingroup bias at age 7–8 (probit regression, $P < 0.001$, $z = 3.58$, $b = 105$), with very little willingness to share with an outgroup member (only 12% of the children shared in the outgroup condition at this age).

The prevalence of egalitarian choices in the envy game developed earlier in the outgroup condition, where the children at age 5–6 already overwhelmingly favour the (1,1) allocation (Fig. 3c). The willingness to remove disadvantageous inequality towards ingroup members becomes so prevalent at age 7–8 that an ingroup–outgroup gap no longer exists. However, averaging across gender hides an important gender effect in the envy game: boys show much stronger parochial tendencies than girls do because boys seem to be much less averse against disadvantageous inequality if the partner is an ingroup member (Fig. 4a; outgroup dummy in probit regression controlling for age, $P = 0.001$, $z = 3.23$, $n = 102$). In contrast, girls do not differentiate in their choices between ingroup and outgroup partners (Fig. 4b; outgroup dummy in probit regression controlling for age, $P = 0.663$, $z = -0.44$, $n = 127$), but, like boys, they also show an increasing trend towards egalitarian choices in the envy game as they become older.

Birth order and sibling effects

We find a strong 'only child' and 'youngest child' effect in the sharing game. Children without siblings showed much more costly sharing behaviour than children with siblings. On average, children without siblings were 28% more likely to share than children with siblings—a highly significant difference (probit regression, $P = 0.006$, $z = 2.75$, $n = 197$) that also exists if we control for income effects (see Supplementary Data). With increasing age, however, the difference between children with and without siblings decreases slightly, as indicated by an interaction effect between 'child without siblings' and 'age' in a probit regression ($P = 0.022$, $z = -2.29$, $n = 197$). Among the children with siblings, we found that—regardless of age—the youngest children in a family were 17% less willing to share than children with younger siblings (probit regression controlling for age, $P = 0.007$, $z = 2.71$, $n = 172$). Thus, it seems that the mere existence of siblings or birth order may have an important role in determining altruistic behaviours. Further analyses and interpretation of the effects of birth order and other demographic and psychological characteristics can be found in the Supplementary Information.

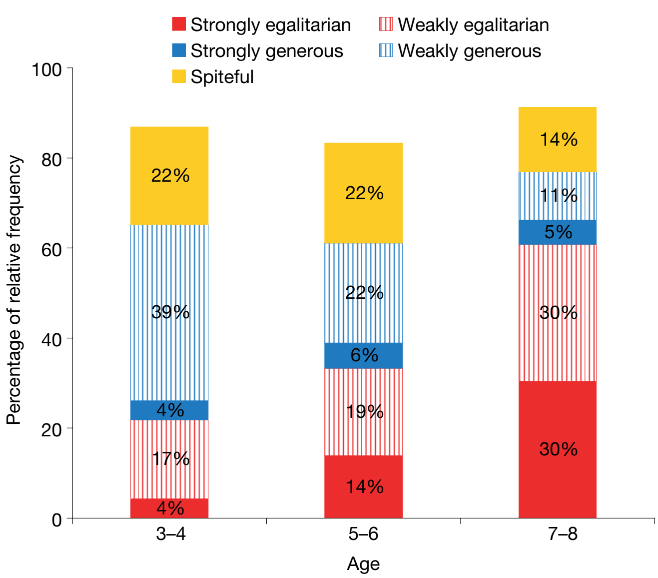


Figure 2 | Behavioural types in the ingroup condition. The figure classifies subjects according to their behaviour in all three games, that is, in the prosocial game ((1,1) versus (1,0)), the sharing game ((1,1) versus (2,0)), and the envy game ((1,1) versus (1,2)). Strongly egalitarian subjects choose the egalitarian allocation in all three games. Weakly egalitarian subjects choose the egalitarian allocation only in the prosocial and the envy games, but not in the sharing game, where egalitarian behaviour is costly. Strongly generous subjects choose the allocation that maximizes the partner's payoff in all three games. Weakly generous subjects maximize the partner's payoff only in the prosocial and the envy games, but not in the sharing game, where generous behaviour is costly. Spiteful subjects choose the allocation that minimizes the partner's payoff in all three games. The percentage of egalitarian subjects increases steeply with age, whereas the share of generous subjects declines. Moreover, most subjects who are willing to share at age 7–8 belong to the egalitarian and not to the generous type of subjects.

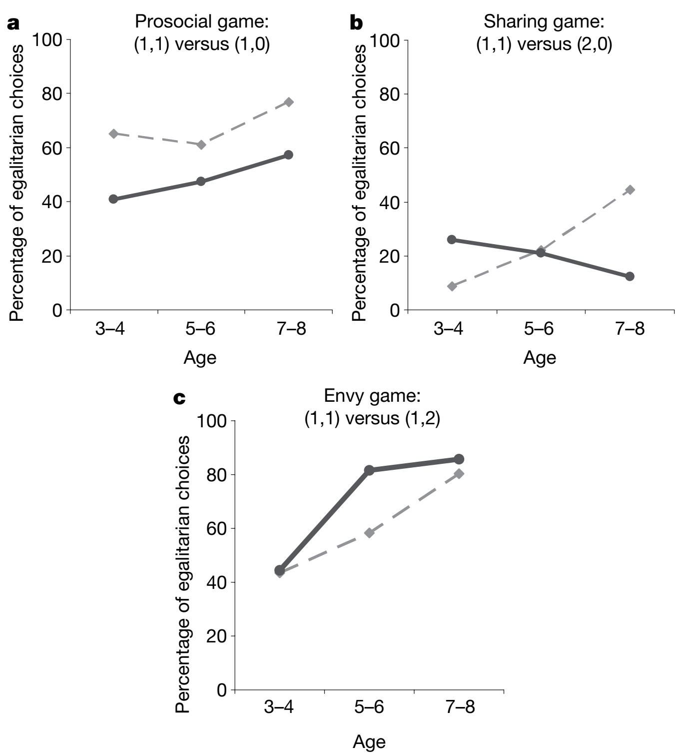


Figure 3 | Egalitarian choices across ingroup and outgroup conditions. Grey dashed lines denote when the partner was an ingroup member; black solid lines represent when the partner was an outgroup member (**a**–**c**). **a**, In the prosocial game, the children remove inequality that favours themselves more often if the partner is an ingroup member. **b**, Egalitarian choices slightly decrease over time in the sharing game (with very little sharing at age 7–8) if the partner is an outgroup member, whereas sharing with ingroup members strongly increases with age, providing strong evidence for parochial altruism in children. **c**, In the envy game, children develop a preference for equality much earlier if the partner is an outgroup member but eventually the aversion against disadvantageous inequality with regard to ingroup members becomes so strong that ingroup–outgroup differences are small.

Discussion

The development of inequality aversion relatively early in childhood is particularly interesting in the light of ethnographic evidence that suggests that there is a strong role of egalitarian 'instincts' in human evolutionary history. There is considerable ethnographic evidence that egalitarian concerns have shaped many human small-scale societies[43,44]. For example, sharing of large hunted game across families seems to have been the rule rather than the exception in small-scale societies, and egalitarian sentiments also play a part in contemporary large-scale societies[4,5]. The important role of egalitarian sentiments in human evolutionary history raises the possibility that there may have been cultural or even genetic transmission that favours egalitarian behaviours. In fact, recent evidence from behavioural genetics suggests that egalitarian behaviour in the ultimatum game has a genetic component[45].

The simultaneous development of altruistic behaviour and parochialism and the gender differences in parochialism are also interesting in view of evolutionary theories that predict that the same evolutionary process[40] jointly determines human altruism and parochialism, meaning that these traits co-evolve in such a way that either both or neither of them evolves. According to the theory, the driving force behind this evolutionary process is frequent intergroup conflict. Because mainly males were involved in intergroup fights, it seems possible that evolution favoured a gender bias in parochialism.

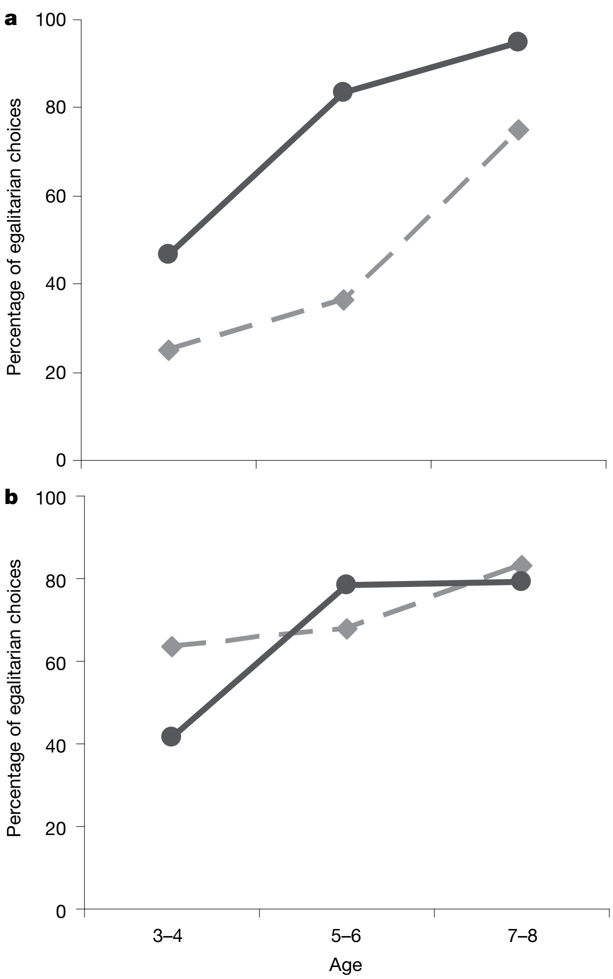


Figure 4 | Gender differences in parochial egalitarianism in the envy game where the child could choose between (1,1) and (1,2). a, b, The propensity of boys (**a**) and girls (**b**) for egalitarian choices in the ingroup (grey dashed line) and outgroup (black solid line) condition. The difference between ingroup and outgroup condition is large for boys but virtually absent for girls.

In fact, a payoff advantage relative to the outgroup may have been particularly advantageous for males because it strengthened the ingroup's position in intergroup conflicts. Males bore the main cost of intergroup conflict in terms of injuries and deaths and often gained more than females in the case of victory because of the increase in the pool of potential mating partners[46]. Thus, evolution may have favoured a greater sensitivity in males for payoff advantages relative to outgroup members. In view of this prediction, we find it notable that boys showed a much stronger ingroup bias than girls in the envy game.

These potential evolutionary roots of human egalitarianism and parochialism do not preclude culture and socialization from playing an important part in other-regarding preferences; they may even be a main factor in their evolution[2,47]. As the children move from an informal playgroup to kindergarten and then on to formal schools, they may learn that equality is a rule that the authorities (for example, the teachers) endorse. Thus, the children probably acquire some of the normative rules of the society which surrounds them during the age period on which we focused[30]. In this context, it is interesting that the motive to increase the partner's payoff declined strongly between ages 3 and 8, whereas egalitarian behaviour strongly increased in this age period (Fig. 2), providing a hint about the content of the normative rule they acquired. The children may also become more sensitive with regard to the opinions of others about themselves, a cognition which requires the ability to understand that one's actions affect what other people believe about oneself. Recent evidence indicates that adult humans care even about what anonymous others may think about them[48,49]. Theory-of-mind and perspective-taking abilities are certainly conducive to such social cognitions. Therefore, if older children care about what anonymous others think about them, they may be more prone to behave in a normatively appropriate way because—owing to their age—they are more likely to have theory-of-mind and perspective-taking abilities[30,50].

Our results also indicate that important inter-species differences occur in other-regarding preferences when compared to the patterns observed in chimpanzees[23,24,27] and marmosets[26]. In the prosocial game, adult chimpanzees, who could allocate food in a face-to-face interaction to a familiar recipient, did not show a significant preference for the (1,1) choice[23,24,27], whereas most of the children develop a preference for the egalitarian choice in this game even though their partner is anonymous. Furthermore, as the sharing game indicates, many children at age 7–8 are also willing to share with an anonymous ingroup member. The facts in the envy game also contrast sharply with experimental findings in both chimpanzees and marmosets. The children developed a widespread aversion against disadvantageous inequality, whereas chimpanzees have been shown to be unwilling to take actions to remove inequality between themselves and a conspecific[24], and marmosets even tend to take actions that generate disadvantageous inequality for themselves if the action provides food for the partner[26]. In view of the decisive role of other-regarding preferences for a species ability to achieve large-scale cooperation, the observed inter-species differences are probably important in explaining why humans are so exceptionally cooperative.

METHODS SUMMARY

Each child played all three games against anonymous partners. To avoid satiation effects, three different currencies (smarties, jellybabies and fizzers) were used in randomized order in the three games. In each game, subjects had two mutually exclusive choices that were represented with two cardboards; we drew two circles with arrows on each (see Supplementary Fig. 1). One arrow pointed to the decision maker, illustrating that the sweets in that circle goes to him or her, whereas the other arrow pointed to a group photo which had been made earlier. In the ingroup condition, the group photo showed the members of the child's playgroup, kindergarten or school, whereas it showed the members of a different playgroup, kindergarten or school in the outgroup condition. The photos were used to communicate the partner's ingroup or outgroup status to the children in the game. Depending on whether the ingroup or the outgroup condition applied, the decision maker was told that the sweet(s) in the other circle (if there were any

in that circle) would be given to one of the children on the ingroup photo or outgroup photo, respectively. If there was no sweet in the circle, it was made clear to the child that the choice of the corresponding cardboard indicates that the partner would receive nothing. We also controlled for the spatial assignment of the different alternative. In the prosocial game, for example (shown in Supplementary Fig. 1), the alternative (1,1) was randomly assigned to the left or right side. Thus, preferences for left or right cannot explain any of our findings. We also made it clear to the children that neither other children nor their parents or teachers will be informed about their decisions (see Supplementary Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 March; accepted 6 June 2008.

1. Bowles, S. *Microeconomics: Behavior, Institutions and Evolution* 93–126 (Princeton Univ. Press, 2004).
2. Boyd, R. T. & Richerson, P. *The Origin and Evolution of Cultures* (Oxford Univ. Press, 2005).
3. Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114,** 817–868 (1999).
4. Fowler, J. H., Johnson, T. & Smirnov, O. Human behaviour: Egalitarian motive and altruistic punishment. *Nature* **433,** E1 (2007).
5. Dawes, C. T. *et al.* Egalitarian motives in humans. *Nature* **446,** 794–796 (2007).
6. Falk, A. & Fischbacher, U. A theory of reciprocity. *Games Econ. Behav.* **54,** 293–315 (2006).
7. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415,** 137–140 (2002).
8. Gürerk, O., Irlenbusch, B. & Rockenbach, B. The competitive advantage of sanctioning institutions. *Science* **312,** 108–111 (2006).
9. Hauert, C. *et al.* Via freedom to coercion: The emergence of costly punishment. *Science* **316,** 1905–1907 (2007).
10. Fong, C., Bowles, S. & Gintis, H. in *Handbook on the Economics of Giving, Reciprocity, and Altruism* (eds Kolm, S.-C. & Ythier, J. M.) (Elsevier, 2006).
11. Milinski, M., Semmann, D. & Krambeck, H. J. Reputation helps solve the 'tragedy of the commons'. *Nature* **415,** 424–426 (2002).
12. Brown, M., Falk, A. & Fehr, E. Relational contracts and the nature of market interactions. *Econometrica* **72,** 747–780 (2004).
13. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity. *Nature* **437,** 1291–1298 (2005).
14. Rockenbach, B. & Milinski, M. The efficient interaction of indirect reciprocity and costly punishment. *Nature* **444,** 718–723 (2006).
15. Harbaugh, W. T. & Krause, K. Children's altruism in public good and dictator experiments. *Econ. Inq.* **38,** 95–109 (2000).
16. de Quervain, D. J.-F. *et al.* The neural basis of altruistic punishment. *Science* **305,** 1254–1258 (2004).
17. Sutter, M. Outcomes versus intentions: On the nature of fair behavior and its development with age. *J. Econ. Psychol.* **28,** 69–78 (2007).
18. Sutter, M. & Kocher, M. G. Trust and trustworthiness across different age groups. *Games Econ. Behav.* **59,** 364–382 (2007).
19. Benenson, J. F., Pascoe, J. & Radmore, N. Children's altruistic behavior in the dictator game. *Evol. Hum. Behav.* **28,** 168–175 (2007).
20. Fliessbach, K. *et al.* Social comparison affects reward-related brain activity in the human ventral striatum. *Science* **318,** 1305–1308 (2007).
21. Cesarini, D. *et al.* Heritability of cooperative behavior in the trust game. *Proc. Natl Acad. Sci. USA* **105,** 3721–3726 (2008).
22. Ellingsen, T. & Johannesson, M. Anticipated verbal feedback induces altruistic behavior. *Evol. Hum. Behav.* **29,** 100–105 (2008).
23. Silk, J. B. *et al.* Chimpanzees are indifferent to the welfare of unrelated group members. *Nature* **437,** 1357–1359 (2005).
24. Jensen, K., Hare, B., Call, J. & Tomasello, M. What's in it for me? Self-regard precludes altruism and spite in chimpanzees. *Proc. Biol. Sci.* **273,** 1013–1021 (2006).
25. Jensen, K., Call, J. & Tomasello, M. Chimpanzees are rational maximizers in an ultimatum game. *Science* **318,** 107–109 (2007).
26. Burkart, J. M., Fehr, E., Efferson, C. & van Schaik, C. P. Other-regarding preferences in a non-human primate: Common marmosets provision food altruistically. *Proc. Natl Acad. Sci. USA* **104,** 19762–19766 (2007).
27. Vonk, J. *et al.* Chimpanzees do not take advantage of very low cost opportunities to deliver food to unrelated group members. *Anim. Behav.* **75,** 1757–1770 (2008).
28. Turiel, E. in *Handbook of Child Psychology* Vol. 3 (eds Damon, W. & Eisenberg, N.) 863–932 (Wiley and Sons, 1998).
29. Knight, G. P. & Kagan, S. Development of prosocial and competitive behaviors in anglo-american and mexican-american children. *Child Dev.* **48,** 1385–1394 (1977).
30. Eisenberg, N. & Mussen, P. H. *The Roots of Prosocial Behavior in Children* (Cambridge Univ. Press, 1989).
31. Thompson, C., Barresi, J. & Moore, C. The development of future-oriented prudence and altruism in preschoolers. *Cogn. Dev.* **12,** 199–212 (1997).
32. Eisenberg, N. & Fabes, R. A. in *Handbook of Child Psychology* Vol. 3 (eds Damon, W. & Eisenberg, N.) 701–778 (Wiley and Sons, 1998).
33. Murnigham, J. K. Ultimatum bargaining by children and adults. *J. Econ. Psychol.* **19,** 415–445 (1998).
34. Warneken, F. & Tomasello, M. Altruistic helping in human infants and young chimpanzees. *Science* **311,** 1301–1303 (2006).
35. Warneken, F. *et al.* Spontaneous altruism by chimpanzees and young children. *PLoS Biol.* **5,** e184 (2007).
36. Charness, G. & Rabin, M. Understanding social preferences with simple tests. *Q. J. Econ.* **117,** 817–869 (2002).
37. Goette, L., Huffman, D. & Meier, S. The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *Am. Econ. Rev.* **96,** 212–216 (2006).
38. Bernhard, H., Fischbacher, U. & Fehr, E. Parochial altruism in humans. *Nature* **442,** 912–915 (2006).
39. Shinada, M., Yamagishi, T. & Ohmura, Y. False friends are worse than bitter enemies: ''Altruistic'' punishment of in-group members. *Evol. Hum. Behav.* **25,** 379–393 (2004).
40. Choi, J. K. & Bowles, S. The coevolution of parochial altruism and war. *Science* **318,** 636–640 (2007).
41. Van Lange, P. A. M. The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *J. Pers. Soc. Psychol.* **77,** 337–349 (1999).
42. Falk, A., Fehr, E. & Fischbacher, U. Driving forces behind informal sanctions. *Econometrica* **73,** 2017–2030 (2005).
43. Boehm, C. Egalitarian behavior and reverse dominance hierachy. *Curr. Anthropol.* **34,** 227–254 (1993).
44. Boehm, C. Impact of the human egalitarian syndrome on darwinian selection mechanics. *Am. Nat.* **150,** S100–S121 (1997).
45. Wallace, B., Cesarini, D., Lichtenstein, P. & Johannesson, M. Heritability of ultimatum game responder behavior. *Proc. Natl Acad. Sci. USA* **104,** 15631–15634 (2007).
46. Chagnon, N. A. Life histories, blood revenge, and warfare in a tribal population. *Science* **239,** 985–992 (1988).
47. Henrich, J. *et al.* ''Economic man'' in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* **28,** 795–855 (2005).
48. Dana, J., Cain, D. M. & Dawes, R. M. What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organ. Behav. Hum. Decis. Process.* **100,** 193–201 (2006).
49. Ellingsen, T. & Johannesson, M. Paying respect. *J. Econ. Perspect.* **21,** 135–149 (2007).
50. Wellman, H. M., Cross, D. & Watson, J. Meta-analysis of theory-of-mind development: The truth about false belief. *Child Dev.* **72,** 655–684 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This paper is part of the Research Priority Program 'Foundations of Human Social Behaviour—Altruism versus Egoism' at the University of Zurich and of the Swiss National Competence Center in research on affective sciences, which is financed by the Swiss National Science Foundation. We also thank N. Kessler for her research assistance during the conduct of the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.F. (efehr@iew.uzh.ch) or B.R. (bettina.rockenbach@uni-erfurt.de).

METHODS

Subjects. We recruited young children (3–8-yr-old) from playschools, kindergartens and first grades for our experiments. We received permission for the experiments from the school president, the school board, the teachers and the parents. Of the parents that we asked, 92% agreed so that a total of 229 children participated in the study: 62 children from playschools, 75 from kindergartens and 92 children from grade one. The children live in Rapperswil-Jona, a small city in the canton of St Gallen, Switzerland. Two women aged between 20 and 25 conducted the experiment. The experimental instructions are reproduced in the Supplementary Information.

Experimental procedures. Each child played the prosocial game, the envy game and the sharing game against anonymous partners. The order of the games was counterbalanced across subjects. The partners came either from the ingroup (same playschool, kindergarten or school) or from an outgroup (a different playschool, kindergarten or school).

Payoffs. We needed an experimental currency desirable for the younger and older children in our sample. Therefore we decided to use various sweets. To avoid satiation effects, a different currency was used in each of the three games (smarties, jellybabies or fizzers), and these were randomized across games. Before the experiment started, we asked each child whether he or she likes those sweets (all of them did). At the end of all three treatments we asked each child whether they liked all three sweets the same or whether they liked one more than the other. If they indicated that they liked some of the sweets more than the others we asked which they liked the most and least (see Supplementary Information). The average 'liking rates' of the different currencies were identical. Moreover, we asked the parents to rate on a 7-point scale how much their children value the sweets and their answers showed that our currencies provide equally strong incentives for the children across all ages.

The choice situation. To ensure that the children could easily understand the choice problem, we used a set up that made the two available choices transparent. The two mutually exclusive choices were represented with two cardboards; we drew two circles with arrows on each (see Supplementary Fig. 1). One arrow pointed to the decision maker, illustrating that the sweet in that circle goes to him or her, whereas the other arrow pointed to a group photo which had been made earlier. In the ingroup condition, the group photo showed the members of the child's playgroup, kindergarten or school, whereas it showed the members of a different playgroup, kindergarten or school in the outgroup condition. The photos were used to communicate the partner's ingroup or outgroup status to the children in the game. Depending on whether the ingroup or the outgroup condition applied, the decision maker was told that the sweet(s) in the other circle (if there were any in that circle) would be given to one of the children on the ingroup photo or outgroup photo, respectively. If there was no sweet in the circle, it was made clear to the child that the choice of the corresponding cardboard implies that the partner would receive nothing. The advantage of this design is that the same procedure can be used for the ingroup and the outgroup condition, and that the children can grasp the partner's ingroup or outgroup status very easily. In fact, many children across all ages immediately expressed their knowledge of the children on the photo, that is, they spontaneously indicated that they knew or did not know the children in the photo. On the basis of this spontaneous insight, it was then easy to explain that the partner will be a member of the child's group or the member of another group.

We also controlled for the spatial assignment of the different alternatives. In the prosocial game, for example, which is shown in Supplementary Fig. 1 (Supplementary Methods), the alternative (1,1) was randomly assigned to be on the left or the right side. Thus, preferences for left or right cannot explain any of our findings. We also made it clear to the children that neither other children nor their parents/teachers would be informed about their decisions.

Before a child played a game we ensured that he or she fully understood the game situation (that is, the available choices, the implications of different choices for the allocation of currencies for 'self' and 'partner', the partner's ingroup–outgroup status, and so on). The subjects had to answer several questions for this purpose (Supplementary Methods). Only three children had problems in answering these questions, which shows that we successfully implemented the three games even for the youngest participants in the sample. The three children who could not answer correctly were excluded from the data analyses (229 is the number of children who correctly answered the questions). Once the children had answered the questions correctly, they were asked to make a decision in the first game before the second game was presented and explained. Because we ensured that the children understood the payoff implications of the two available choices very well and because it was clear that the choices were mutually exclusive (that is, only one cardboard could be chosen), the children did not make mistakes while choosing, for example, by indicating that they wanted to choose both cardboards. Nor did any children reverse their opinion during the choice process, that is, children who first chose one cardboard but later switched to the other.

Questionnaire. After all children had participated in the experiment we sent the parents a questionnaire in which we asked them about characteristics of their child, such as whether he or she can easily imagine how other children feel ('empathy'), whether there are siblings, the birth order, who primarily cares for the child during the day, the number of regular playmates, and so on. We sent 198 questionnaires to the parents and 161 questionnaires were sent back. The questionnaire is reprinted in the Supplementary Information.

Vol 454 | 28 August 2008 | doi:10.1038/nature07268

ARTICLES

Major viral impact on the functioning of benthic deep-sea ecosystems

Roberto Danovaro[1], Antonio Dell'Anno[1], Cinzia Corinaldesi[1], Mirko Magagnini[1], Rachel Noble[2], Christian Tamburini[3] & Markus Weinbauer[4]

Viruses are the most abundant biological organisms of the world's oceans. Viral infections are a substantial source of mortality in a range of organisms—including autotrophic and heterotrophic plankton—but their impact on the deep ocean and benthic biosphere is completely unknown. Here we report that viral production in deep-sea benthic ecosystems worldwide is extremely high, and that viral infections are responsible for the abatement of 80% of prokaryotic heterotrophic production. Virus-induced prokaryotic mortality increases with increasing water depth, and beneath a depth of 1,000 m nearly all of the prokaryotic heterotrophic production is transformed into organic detritus. The viral shunt, releasing on a global scale ~0.37–0.63 gigatonnes of carbon per year, is an essential source of labile organic detritus in the deep-sea ecosystems. This process sustains a high prokaryotic biomass and provides an important contribution to prokaryotic metabolism, allowing the system to cope with the severe organic resource limitation of deep-sea ecosystems. Our results indicate that viruses have an important role in global biogeochemical cycles, in deep-sea metabolism and the overall functioning of the largest ecosystem of our biosphere.

Viruses are by far the most abundant 'life forms' in the world's oceans (approximately 4×10^{30} viruses)[1], exceeding prokaryotic abundance by at least one order of magnitude[1,2]. Increasing evidence indicates that viral infection may be responsible for the high mortality of autotrophic and heterotrophic organisms in surface oceans[3–5], with cascading effects on carbon cycling and nutrient regeneration[6,7]. Viral lysis of infected microbes transforms their cell contents and biomass into organic detritus (both dissolved and particulate), which can then be used again by non-infected prokaryotes (that is, viral shunt)[1,6,8]. This process supports prokaryotic heterotrophic production, but it also decreases the efficiency of the carbon transfer to higher trophic levels[6] and influences the carbon budget of the oceans, thereby modifying the amount of carbon transferred by sinking particles from the surface waters towards the ocean floor[5]. Therefore, the integration of the viral component into trophodynamic and biogeochemical models is of primary importance for an improved understanding of the function of the world's oceans[5].

Deep-sea ecosystems cover about 65% of the Earth's surface and play an important part in biomass production and biogeochemical cycles on a global scale[9]. These processes are largely mediated by benthic prokaryotes, which use organic detritus for biomass production and respiration[10]. Deep-sea ecosystems are dark and extreme environments that lack photosynthetic primary production, depend on the carbon export from the surface oceans and are characterized by severe organic nutrient limitation[9]. The biomass of all faunal components decreases exponentially with increasing water depth[11], but this does not apply to benthic prokaryotic biomass[11], which also has a relatively high metabolic rate[12]. Prokaryotic biomass, in the top 10-cm of deep-sea sediments, is approximately 160 Pg, which is equivalent to 30%–45% of the total microbial carbon on Earth[13]. The huge, nitrogen- and phosphorus-rich, prokaryotic biomass represents a potentially enormous and high quality food source for benthic consumers in deep-sea ecosystems[14], but *in situ* experiments suggest that prokaryotic biomass does not contribute significantly to the food requirements of higher trophic levels[15]. The high prokaryotic biomass in food-limited deep-sea ecosystems and the non-use of this component represent two unsolved paradoxes of the deep oceans.

Viruses in benthic deep-sea ecosystems

We report here that viral abundance in surface sediments (top 1-cm) worldwide is high at all depths, from the shelf-break down to the abyssal sediments (range, $0.10–3.53 \times 10^9\ g^{-1}$; mean, $0.96 \times 10^9\ g^{-1}$ (Supplementary Table 1), which are equivalent to $0.83–28.2 \times 10^{12}$ viruses m^{-2} with a mean of $7.65 \pm 0.32 \times 10^{12}$ viruses m^{-2}; Fig. 1a), with values similar to those reported for coastal sediments[16,17]. A high viral abundance in deep-sea sediments could be due to the supply of virioplankton adsorbed onto sinking particles[18] and/or to the high *in situ* viral production rates[16,17]. We found that the downward flux of viruses associated with settling particles to the deep sea was extremely low (Supplementary Table 2), whereas direct measurements, on the basis of different independent approaches, indicate that viral production rates in deep-sea sediments are consistently high, and that viruses are produced *in situ* (Fig. 1b and Supplementary Figs 2 and 3).

Further experiments conducted in both *in situ* pressure and decompressed samples indicated that >99% of the prokaryotic infections are lytic, both in deep-sea sediments and at the benthic boundary layer (Supplementary Fig. 4). Deep-benthic viruses are therefore functionally active and are capable of killing the infected prokaryotes. We also estimated the burst size of prokaryotes (defined as the number of viruses that cause the lysis of the host cell) in deep-sea sediments using different independent approaches, which ranged from 11 to 106 (mean of time course experiments, 45; median, 38; Supplementary Table 3 and Supplementary Fig. 5).

[1]Department of Marine Science, Faculty of Science, Polytechnic University of Marche, Via Brecce Bianche, 60131 Ancona, Italy. [2]Institute of Marine Sciences, The University of North Carolina at Chapel Hill, 3431 Arendell Street, Morehead City, North Carolina 28557, USA. [3]Université de la Méditerranée, Centre d'Océanologie de Marseille, UMR 6117–CNRS, Campus de Luminy, Case 901, 163 Avenue de Luminy, 13288 Marseille, Cedex 9, France. [4]CNRS; Microbial Ecology & Biogeochemistry Group, Laboratoire d'Océanographie de Villefranche, 06234 Villefranche-sur-Mer, France, Université Pierre et Marie Curie-Paris 6, Laboratoire d'Océanographie de Villefranche, 06234 Villefranche-sur-Mer, France.

Changes in hydrostatic pressure can alter the physiological status of the prokaryotes[19–21], thus potentially influencing viral production measurements. Analyses of viral and prokaryotic production in deep-sea sediment samples at *in situ*, decompressed (at 0.1 MPa) and re-pressurized conditions (see Supplementary Methods for details) demonstrated that production rates are not significantly influenced by deep-sea sample recovery (Supplementary Figs 6 and 7), allowing the conclusion that the high viral production rates are not due to artefacts.

Combining values of viral production and viral abundance in deep-sea sediments, we conclude that deep-sea viriobenthos is a highly dynamic and active component of deep-sea ecosystems, with a viral turnover (range, 0.15–8.2 d^{-1}, mean, 1.7 ± 0.1 d^{-1}) comparable to, or higher than, that reported for highly productive aquatic ecosystems[16].

Virus-induced prokaryotic mortality

Viral abundance and production did not change significantly with water depth, even when latitude, longitude and sampling time were included as covariates in the statistical analyses (Supplementary Table 4). However, viral and prokaryotic abundance (Fig. 1a, c) covaried across a wide range of depths and environmental conditions (Fig. 2a). Also, viral production and prokaryotic heterotrophic production (Fig. 1b, d) were significantly and positively related (Fig. 2b). These positive relationships suggest a dependence of viral replication on the host abundance and metabolism, and indicate that the observed variability across depths is not due to a methodological artefact. Thus, our data confirm the emerging view that deep-sea benthic ecosystems are more dynamic than previously thought[14] and suggest that the extreme conditions of the deep oceans are not hostile for viral development.

We estimated the impact of viruses on benthic prokaryotic production, calculating the number of prokaryotes killed as the ratio between viral production and prokaryotic burst size. Using the mean burst size of the time course experiments (burst size of 45) we calculated that viruses are responsible for the abatement of 80% of the total

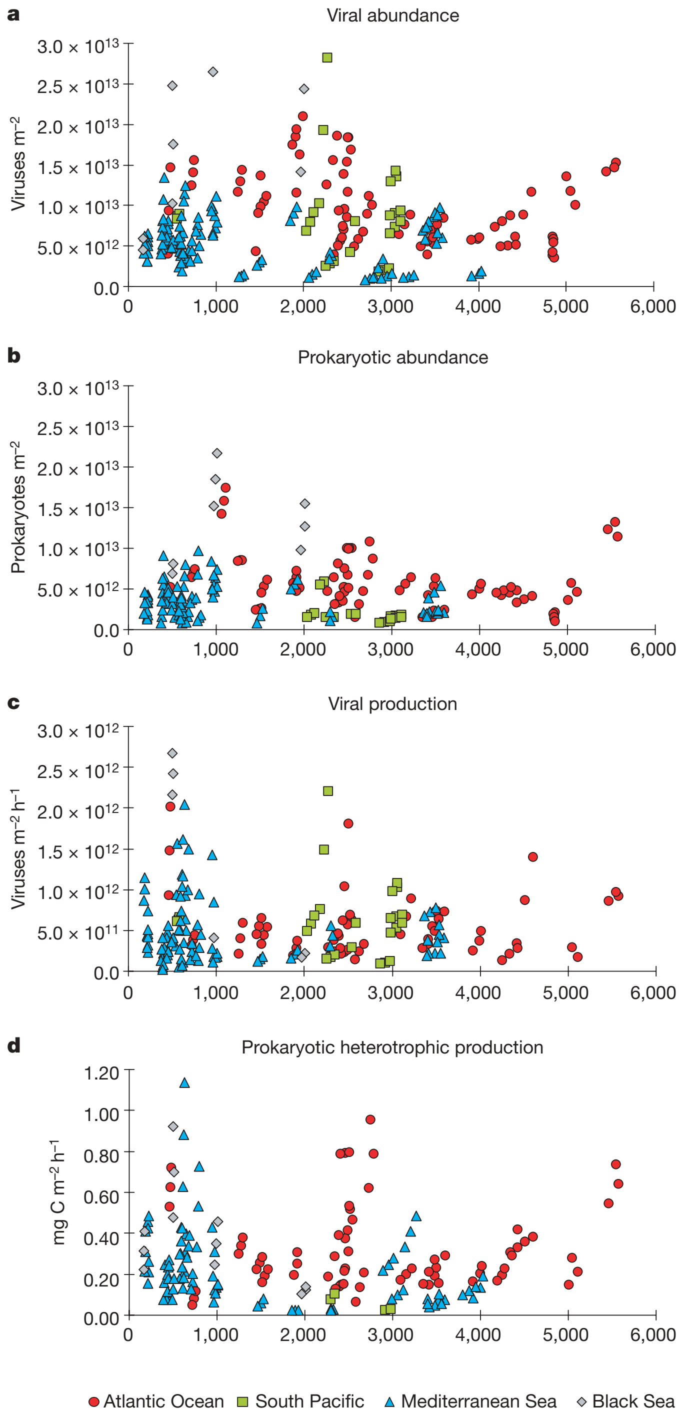


Figure 1 | Bathymetric patterns of viruses and prokaryotes in deep-sea sediments worldwide. Reported are the viral abundance (**a**; expressed as the number of viruses per m^2), the prokaryotic abundance (**b**; expressed as the number of prokaryotes per m^2), the viral production (**c**; expressed as the number of viruses produced per m^2 per h), and the prokaryotic heterotrophic production (**d**; expressed as mg C m^{-2} h^{-1}). Data originate from the Atlantic Ocean (temperate systems), the South Pacific Ocean (tropical warm systems), the western and eastern Mediterranean Sea (temperate warm systems) and the Black Sea (anoxic system).

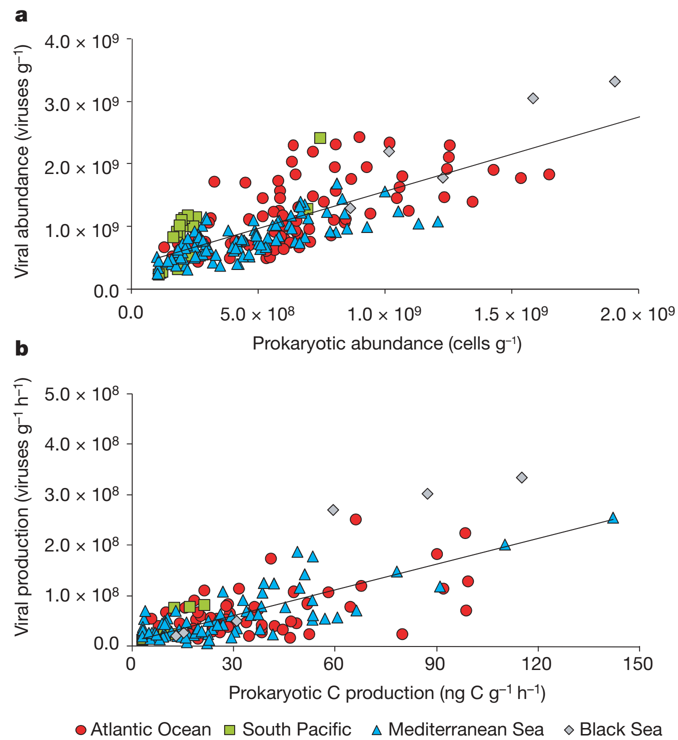


Figure 2 | Relationships between viruses and prokaryotes in deep-sea sediments worldwide. Reported are the relationship between viral and prokaryotic abundance (**a**; regression equation: $y = 1.19x + 3.63 \times 10^8$; $n = 206$, $R^2 = 0.56$, $P < 0.001$), and the relationship between viral and prokaryotic production (**b**; regression equation: $y = 1.71 \times 10^6 x + 9.89 \times 10^6$; $n = 186$, $R^2 = 0.57$, $P < 0.001$). Data originate from the Atlantic Ocean (temperate systems), the South Pacific Ocean (tropical warm systems), the western and eastern Mediterranean Sea (temperate warm systems) and the Black Sea (anoxic system).

prokaryotic heterotrophic production in global deep-sea sediments. Virus-induced prokaryotic mortality increases with water depth from 16% ± 3% (mean ± s.e.m., $n = 11$) in coastal sediments to 64% ± 3% in mesopelagic sediments (>160–1,000-m depth, $n = 41$) to 89% ± 2% in sediments beneath 1,000-m depth ($n = 67$; Fig. 3). These are conservative estimates, as the use of the mean or median burst sizes obtained from the three independent approaches (burst sizes of 37 or 33, respectively; Supplementary Table 3), or the values of burst sizes frequently used in the literature (20–30)[2,7], would result in the abatement of ~90% or more of the global prokaryotic carbon production.

These results indicate that viruses are the main agents of mortality of prokaryotes in deep-sea sediments worldwide, and determine the shunt of most of the prokaryotic carbon production into organic detritus. This finding has important implications for the comprehension of food webs and biogeochemical processes in deep-sea ecosystems. Viruses, killing most of the prokaryotes produced, drastically reduce the microbial biomass potentially available to higher trophic levels in deep-sea sediments. These results help to explain the paradox of deep-sea ecosystems in which prokaryotic biomass is largely unused by deep-sea fauna, despite severe food limitation[15].

Implications on ecosystem functioning

The amount of organic resources in the deep sea is typically extremely low (organic carbon concentrations are 10–20 times lower than in coastal systems), mostly composed of refractory compounds[22] and could represent a limiting factor for prokaryotic metabolism and turnover[23,24]. There is evidence to suggest that intracellular material released by viral lysis can be an important and readily bioavailable organic source for prokaryotes[25,26]. The significant relationship between the release of carbon from prokaryotic lysis and prokaryotic turnover in deep-sea sediments reported here (Fig. 4) suggests that a stronger viral shunt is associated with higher prokaryotic growth rates. This relationship allows us to speculate that the viral shunt, on the one hand, kills an important fraction of benthic prokaryotes and reduces the competition for available resources and, on the other, generates labile, highly bioavailable material and stimulates the metabolism of uninfected deep-sea prokaryotes. These results contribute to explaining the paradox of a fast prokaryotic turnover in food-limited, deep-sea ecosystems.

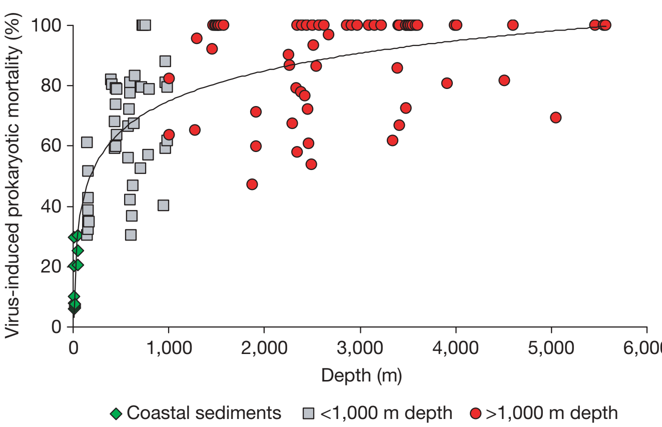


Figure 3 | Bathymetric patterns of virus-induced prokaryotic mortality. The virus-induced prokaryotic mortality was calculated as the number of cells killed divided by the total number of prokaryotes produced per hour, and expressed as a percentage. The regression equation is $y = 14.46\ln(x) - 24.98$; $n = 119$, $R^2 = 0.686$, $P < 0.001$. Data from coastal sediments (green diamonds) are derived from the literature[17,26,27]. Data for mesopelagic (grey squares), and bathyal and abyssal sediments (red circles) originate from the Atlantic Ocean (temperate systems), the South Pacific Ocean (tropical warm systems), the western and eastern Mediterranean Sea (temperate warm systems) and the Black Sea (anoxic system).

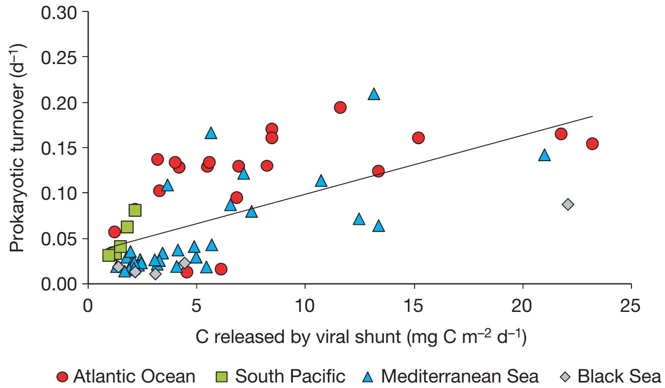


Figure 4 | Relationship between carbon released by viral lysis of prokaryotic biomass (viral shunt) and by prokaryotic turnover. Prokaryotic turnover was expressed per day. The regression equation is: $y = 0.0065x + 0.034$; $n = 63$, $R^2 = 0.391$, $P < 0.01$. Data originate from the Atlantic Ocean (temperate systems), the South Pacific Ocean (tropical warm systems), the western and eastern Mediterranean Sea (temperate warm systems) and the deep Black Sea (anoxic system).

The release of labile organic material from killed prokaryotic cells has important implications for carbon cycling and nutrient regeneration on a global scale. We estimate that the viral shunt, on a global scale, can release from ~0.37 Gt C yr^{-1} (the most conservative estimate on the basis of the highest value of burst size) to 0.63 Gt C yr^{-1} (if a burst size of 45 is used). Therefore, the viral shunt is an essential source of labile organic compounds to deep-sea ecosystems. Using conservative conversion factors[14], we estimate that the organic carbon supplied by viral shunt contributes to around 35% of the total benthic prokaryotic metabolism, increases their turnover and promotes the recycling of key elements (including nitrogen and phosphorus associated with prokaryotic biomass). Given the global relevance of biogeochemical processes occurring in deep-benthic ecosystems[10], our results indicate that the deep-sea viral shunt is a crucial process, so far neglected, which has to be included in global models of carbon, nitrogen and phosphorus cycling and nutrient flows.

We conclude that viral infection has an important part in the functioning of the largest ecosystem of the biosphere by controlling benthic prokaryotic biomass (top down, predatory control), by stimulating prokaryotic metabolism (bottom-up mechanism), and by accelerating biogeochemical processes.

METHODS SUMMARY

We collected a data set of 232 deep-sea sediment samples and measured the impact of viruses on deep-sea benthic prokaryotes and biogeochemical cycles. Our data cover latitudes from 79° N to 34° S, and all depths from 165 m to 5,571 m (sampling location and details on the data set is provided in Supplementary Table 1 and Fig. 1), and include deep-sea sites spanning a wide range of bottom-water temperatures and trophic conditions. Measurements of viral and prokaryotic abundance and production were carried out synoptically on the same samples and by using the same protocols for the entire data set, thus ensuring methodological consistency.

Viral and prokaryotic abundances in deep-sea sediments were determined by epifluorescence microscopy using highly sensitive fluorochromes. For the determination of viral production, two different procedures were applied and compared. The first is based on a dilution technique of samples with virus-free sea water, and allows the determination of viral production on the basis of the increase in viral number over time. The second is based on the use of ^{3}H-thymidine, and allows the determination of viral production through the assessment of the incorporation rates of radiolabelled substrates into viral genomes. Radiolabelled substrates were also used to determine prokaryotic heterotrophic carbon production. To substantiate the measurements of viral and prokaryotic heterotrophic carbon production carried out at 0.1 MPa, further experiments were carried out under *in situ* pressure conditions using *ad hoc* high pressure sampling devices. The ratio between viral production and prokaryotic burst size (estimated using three independent approaches) was used to estimate the

prokaryotic mortality induced by viruses. Finally, the relative importance of lysogenic infections was investigated by time course experiments of deep-sea samples previously treated with specific agents inducing the lytic cycle.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 May; accepted 15 July 2008.

1. Suttle, C. A. Viruses in the sea. *Nature* 437, 356–361 (2005).
2. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28, 127–181 (2004).
3. Proctor, L. M. & Fuhrman, J. A. Viral mortality of marine bacteria and cyanobacteria. *Nature* 343, 60–62 (1990).
4. Suttle, C. A., Chan, A. M. & Cottrell, M. T. Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* 347, 467–469 (1990).
5. Suttle, C. A. Marine viruses—major players in the global ecosystem. *Nature Rev. Microbiol.* 5, 801–812 (2007).
6. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548 (1999).
7. Wommack, K. E. & Colwell, R. R. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64, 69–114 (2000).
8. Wilhelm, S. W. & Suttle, C. A. Viruses and nutrient cycles in the sea. *Bioscience* 49, 781–788 (1999).
9. Gage, J. D. & Tyler, P. A. *Deep-Sea Biology: A Natural History of Organisms at the Deep Sea Floor* (Cambridge Univ. Press, 1991).
10. Jahnke, R. A. The global ocean flux of particulate organic carbon: areal distribution and magnitude. *Glob. Biogeochem. Cycles* 10, 71–88 (1996).
11. Rex, M. A. *et al.* Global bathymetric patterns of standing stock and body size in the deep-sea benthos. *Mar. Ecol. Prog. Ser.* 317, 1–8 (2006).
12. Yayanos, A. A. Microbiology to 10,500 meters in the deep sea. *Annu. Rev. Microbiol.* 49, 777–805 (1995).
13. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: The unseen majority. *Proc. Natl Acad. Sci. USA* 95, 6578–6583 (1998).
14. Dell'Anno, A. & Danovaro, R. Extracellular DNA plays a key role in deep-sea ecosystem functioning. *Science* 309, 2179 (2005).
15. Witte, U. *et al. In situ* experimental evidence of the fate of a phytodetritus pulse at the abyssal sea floor. *Nature* 424, 763–766 (2003).
16. Danovaro, R. *et al.* Viriobenthos in freshwater and marine sediments: a review. *Freshwat. Biol.* 53, 1186–1213 (2008).
17. Siem-Jørgensen, M., Glud, R. N. & Middelboe, M. Viral dynamics in a coastal sediment: seasonal pattern, controlling factors and relations to the pelagic-benthic coupling. *Mar. Biol. Res.* 4, 165–179 (2008).
18. Mari, X., Kerros, M.-E. & Weinbauer, M. G. Virus attachment to transparent exopolymeric particles along trophic gradients in the Southwestern Lagoon of New Caledonia. *Appl. Environ. Microbiol.* 73, 5245–5252 (2007).
19. Tamburini, C., Garcin, J., Ragot, M. & Bianchi, A. Biopolymer hydrolysis and bacterial production under ambient hydrostatic pressure through a 2000 m water column in the NW Mediterranean. *Deep-Sea Res. II* 49, 2109–2123 (2002).
20. Tamburini, C., Garcin, J. & Bianchi, A. Role of deep-sea bacteria in organic matter mineralization and adaptation to hydrostatic pressure conditions in the NW Mediterranean Sea. *Aquat. Microb. Ecol.* 32, 209–218 (2003).
21. Bartlett, D. H. Pressure effects on *in vivo* microbial processes. *Biochim. Biophys. Acta* 1595, 367–381 (2002).
22. Druffel, E. R. M. & Robinson, D. H. Is the deep sea on a diet? *Science* 284, 1139–1140 (1999).
23. Deming, J. W. & Barross, J. A. in *Organic Geochemistry: Principles and Applications* (eds Engel, M. H. & Macko, S.A.) 119–144 (Plenum, 1993).
24. Danovaro, R., Della Croce, N., Dell'Anno, A. & Pusceddu, A. A depocenter of organic matter at 7800-m depth in the SE Pacific Ocean. *Deep-Sea Res. I* 50, 1411–1420 (2003).
25. Noble, R. T. & Fuhrman, J. A. Breakdown and microbial uptake of marine viruses and other lysis products. *Aquat. Microb. Ecol.* 20, 1–11 (1999).
26. Corinaldesi, C., Dell'Anno, A. & Danovaro, R. Viral infection plays a key role in extracellular DNA dynamics in marine anoxic systems. *Limnol. Oceanogr.* 52, 508–516 (2007).
27. Mei, M. L. & Danovaro, R. Virus production and life strategies in aquatic sediments. *Limnol. Oceanogr.* 49, 459–470 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was financially supported by the EU within the framework of the project HERMES. C.T. was supported by the ANR POTES.

Author Contributions R.D. performed the project planning; A.D., C.C., M.M. and C.T. performed the experimental work; R.D., A.D., C.C., M.M., R.N., C.T. and M.W. performed the data analysis; R.D., A.D., C.C., M.M., R.N., C.T. and M.W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.D. (r.danovaro@univpm.it).

METHODS

Sampling sites and sampling strategy. Sampling was carried out in the Atlantic Ocean (~30% of the investigated sites), the Pacific Ocean (~25%), the Mediterranean Sea (~42%) and the Black Sea (~3%). Overall, 38% of the sampling sites were located between depths of 165 m and 1,000 m, 35% were between depths of 1,000 m and 3,000 m, and 27% were between depths of 3,000 m and 5,571 m. Details of the materials and methods used in the present study are reported in the Supplementary Information.

Viral abundance, production and lysogenic fraction. For the determination of viral abundance, fresh and unpreserved sediments were processed immediately after retrieval to minimize viral loss due to the use of preservatives (for example, formaldehyde)[28–30]. Viral abundances were determined by epifluorescence microscopy after extraction of benthic viruses from the sediments and subsequent staining with SYBR Green I (ref. 29). Viral abundance was normalized to gram of sediment dry weight after desiccation (60 °C, 24 h) and converted in surface units (viruses m^{-2}), assuming a constant value of 1.8 g cm^{-3} for the sediment dry bulk density.

Viral production in deep-sea sediment samples was determined by the dilution technique[27,31] (see Supplementary Methods for details). This technique has been the most widely used so far in benthic environments (thus enabling a comparison with available data) and has the advantage of minimizing the impact of protozoan and/or faunal grazing during the incubations[16]. The reliability of the protocol was tested by comparing the results of viral production obtained using different sediment dilutions (1:2, 1:5, 1:10 (v/v) of sediment versus virus-free sea water; see Supplementary Methods), and by comparing the results of viral production acquired by the dilution technique with those obtained by the incorporation of ^{3}H-thymidine by deep-sea viruses[32,33]. All incubations were carried out in the dark and at *in situ* temperature. The effects of pressure were tested for both techniques (dilution and ^{3}H-thymidine incorporation) by comparing viral production from decompressed samples (0.1 MPa) with samples maintained at *in situ* pressure (Supplementary Figs 3 and 6). To estimate the lysogenic fraction among natural assemblages of deep-sea prokaryotes, we used one of the most widely used induction agents (mitomycin carbon[27,34]; Supplementary Methods). Deep-sea sediment samples were inoculated with mitomycin carbon and incubated at *in situ* pressure and temperature. Decompressed samples (0.1 MPa) were used as a control (Supplementary Fig. 4).

Prokaryotic abundance, biomass and production. Prokaryotic abundance in deep-sea sediments was determined from the same sediment samples used for the viral counts. Prokaryotic cells were extracted from the sediments according to standard procedures, stained with SYBR Green I and counted by epifluorescence microscopy[27–29]. For the determination of the prokaryotic biomass, the cell biovolume was converted into carbon content assuming 310 fg C μm^{-3} as a conversion factor[24,35]. Prokaryotic abundance and biomass were normalized to sediment dry weight after desiccation (60 °C, 24 h). Prokaryotic heterotrophic production in deep-sea sediments was determined by ^{3}H-leucine incorporation (up to 3 h of incubation in the dark at *in situ* temperature)[24]. The effect of pressure on benthic prokaryotic heterotrophic production was tested using ^{3}H-thymidine incorporation[36], synoptically with tests conducted for viral production and compared to decompressed and non-decompressed samples (Supplementary Fig. 7).

Effects of pressure on viral and prokaryotic production. In 2005 and 2007, two series of independent experiments were conducted to test the effects of pressure on different microbial variables (that is, viral production, lysogenic fraction and prokaryotic heterotrophic production). In 2005, 500-ml high-pressure bottles were used for the sampling of the water–sediment interface (collected at 2,474 and 3,526 m depth). Samples were incubated at *in situ* pressure and temperature conditions[19,20] with special attention to maintain the same pressure during the entire experiment (recovery, transfer and incubations). In 2007, sediment samples were collected at 3,500 m depth by means of a multiple corer, retrieved on board, immediately transferred in sterilized syringes and re-pressurized in high-pressure bottles, kept in the dark and under *in situ* temperature conditions for subsequent viral and prokaryotic production measurements (see details in Supplementary Methods). All variables were determined by time course experiments and, at each sampling time, aliquots of the incubated samples were withdrawn directly from the pressurized experimental system (see Supplementary Methods). For all experiments, extra samples were collected, using the same sampling equipments without pressurization for evaluating the effect of incubation at 0.1 MPa on all of the measured microbial variables.

Virus-induced prokaryotic mortality. The number of prokaryotes killed by viruses per gram of sediment per hour was calculated as the ratio between viral production (as the number of benthic viruses produced per gram of sediment per hour) and burst size (as the number of viruses released by cell lysis because of viral infection). Prokaryotic burst size was estimated using three independent approaches: (1) from time course experiments in which the increase of viral abundance with time was divided by the number of prokaryotic cells killed; (2) by applying a regression analysis between cell biovolume and burst size derived from the literature[37]; and (3) using transmission electron microscopy for counting the number of phages in deep-sea benthic prokaryotic cells[38]. Details on the procedures are reported in Supplementary Information. The virus-induced prokaryotic mortality was calculated as the number of cells killed divided by the total number of prokaryotes produced per hour, and expressed as a percentage[27].

Statistical analyses. To test whether the investigated benthic microbial variables were influenced by water depth, latitude, longitude or sampling period, we carried out a multivariate multiple regression analysis. All analyses were done with the routine distance-based multivariate analysis for a linear model (DISTLM) forward, and the effects of depth, sampling time, longitude and latitude were included as covariates in the analyses. *P* values were obtained with 4,999 permutations of residuals under the reduced model[39].

28. Helton, R. R., Liu, L. & Wommack, K. E. Assessment of factors influencing direct enumeration of viruses within estuarine sediments. *Appl. Environ. Microbiol.* **72,** 4767–4774 (2006).
29. Danovaro, R., Dell'Anno, A., Trucco, A., Serresi, M. & Vanucci, S. Determination of virus abundance in marine sediments. *Appl. Environ. Microbiol.* **67,** 1384–1387 (2001).
30. Wen, K., Ortmann, A. C. & Suttle, C. A. Accurate estimation of viral abundance by epifluorescence microscopy. *Appl. Environ. Microbiol.* **70,** 3862–3867 (2004).
31. Wilhelm, S. W., Briden, S. & Suttle, C. A. A dilution technique for the direct measurement of viral production: a comparison in stratified and tidally mixed coastal waters. *Microb. Ecol.* **43,** 168–173 (2002).
32. Steward, G. F., Wikner, J., Smith, D. C., Cochlan, W. P. & Azam, F. Estimation of virus production in the sea: 1. Method development. *Mar. Microb. Food Webs* **6,** 57–78 (1992).
33. Fuhrman, J. A. & Noble, R. T. Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnol. Oceanogr.* **40,** 1236–1242 (1995).
34. Weinbauer, M. G., Brettar, I. & Höfle, M. G. Lysogeny and virus-induced mortality of bacterioplankton in surface, deep, and anoxic marine waters. *Limnol. Oceanogr.* **48,** 1457–1465 (2003).
35. Fry, J. C. in *Methods in Aquatic Bacteriology* (ed. Austin, B.) 27–72 (John Wiley & Sons, 1990).
36. Fuhrman, J. A. & Azam, F. Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters: evaluation and field results. *Mar. Biol.* **66,** 109–120 (1982).
37. Weinbauer, M. G. & Höfle, M. G. Size-specific mortality of lake bacterioplankton by natural virus communities. *Aquat. Microb. Ecol.* **15,** 103–113 (1998).
38. Weinbauer, M. G., Fuks, D. & Peduzzi, P. Distribution of viruses and dissolved DNA along a coastal trophic gradient in the Northern Adriatic Sea. *Appl. Environ. Microbiol.* **59,** 4074–4082 (1993).
39. Anderson, M. J. *DISTLM v.5: a FORTRAN Computer Program to Calculate a Distance-Based Multivariate Analysis for a Linear Model* (Department of Statistics, Univ. Auckland, 2004).

ARTICLES

Misfolded proteins partition between two distinct quality control compartments

Daniel Kaganovich[1], Ron Kopito[1] & Judith Frydman[1]

The accumulation of misfolded proteins in intracellular amyloid inclusions, typical of many neurodegenerative disorders including Huntington's and prion disease, is thought to occur after failure of the cellular protein quality control mechanisms. Here we examine the formation of misfolded protein inclusions in the eukaryotic cytosol of yeast and mammalian cell culture models. We identify two intracellular compartments for the sequestration of misfolded cytosolic proteins. Partition of quality control substrates to either compartment seems to depend on their ubiquitination status and aggregation state. Soluble ubiquitinated misfolded proteins accumulate in a juxtanuclear compartment where proteasomes are concentrated. In contrast, terminally aggregated proteins are sequestered in a perivacuolar inclusion. Notably, disease-associated Huntingtin and prion proteins are preferentially directed to the perivacuolar compartment. Enhancing ubiquitination of a prion protein suffices to promote its delivery to the juxtanuclear inclusion. Our findings provide a framework for understanding the preferential accumulation of amyloidogenic proteins in inclusions linked to human disease.

The strong correlation between the accumulation of aggregated proteins in amyloid inclusions and the onset of several neurodegenerative diseases calls for a better understanding of the mechanisms and functions of inclusion formation. Research indicating that soluble aggregation intermediates have a toxic 'gain of function' activity suggests that regulated formation of protein inclusions serves cytoprotective functions, such as sequestering misfolded species[1–7], and it may also facilitate their clearance[8–12]. It is unknown whether inclusions contain only terminally aggregated proteins or whether they also sequester soluble misfolded conformations[13]. Intriguingly, although all proteins can form amyloid-like inclusions after misfolding[1,14], only a handful of proteins cause amyloidosis and disease[2]. In principle, these amyloidogenic disease-related proteins may interact differently with the cellular quality control machinery. Thus, characterization of the pathways leading to inclusion formation is critical for understanding the basis of protein conformation disorders.

Cellular inclusions form in an organized process that seems to be conserved from yeast to mammalian cells[2,8,15]. Distinct inclusions with specific characteristics have been observed[13,16–19], including insoluble perinuclear inclusions (called aggresomes) that co-localize with the microtubule organizing centre[20], perinuclear inclusions containing soluble endoplasmic reticulum associated protein degradation (ERAD) substrates[18,21], and inclusions co-localizing with autophagic markers[9,10]. It is unclear whether all these observations pertain to the same compartment or what underlies the distinct solubility and long-term fates observed for different quality control substrates in these inclusions.

Unlike amyloidogenic proteins, little is known about the fate of 'normal' misfolded cytosolic globular proteins[22]. Protein misfolding can arise as a consequence of stress-induced denaturation, destabilizing missense mutations or lack of oligomeric assembly partners. To examine how cytosolic quality control proceeds in these different scenarios, we chose a panel of model substrates corresponding to each case (Fig. 1) and compared their fate to that of model amyloidogenic proteins (Fig. 2). Our findings show that the quality control machinery partitions misfolded proteins, on the basis of their ubiquitination state and solubility, among two distinct quality control compartments. Interestingly, amyloidogenic proteins are preferentially sorted to only one of these compartments. These distinct quality control compartments may represent two cellular strategies for the sequestration of aggregation prone, potentially toxic polypeptides.

Two compartments for misfolded cytosolic proteins

To determine the fate of cytosolic misfolded substrates, we initially followed a destabilized Ubc9 variant that misfolds above 30 °C (refs 23, 24; Fig. 1a). Ubc9^{ts}, fused to green fluorescent protein (GFP) to facilitate detection (GFP–Ubc9^{ts}), was expressed under the control of a galactose-regulated promoter. Glucose addition repressed expression, allowing us to follow the fate of GFP–Ubc9^{ts} from the earliest stages of protein misfolding after shift to 37 °C (Fig. 1a). At permissive temperatures, GFP–Ubc9^{ts} was native and diffuse, similar to wild-type GFP–Ubc9 (Fig. 1b, 0 min, compare with wild type panel 120 min). GFP–Ubc9^{ts} misfolding led to degradation by the ubiquitin–proteasome pathway, as reported for untagged Ubc9^{ts} (Fig. 1b, compare 5 min and 60 min; and Fig. 1d, left panel)[23,24]. During degradation we observed transient accumulation of Ubc9^{ts} in distinct cytosolic puncta and inclusions that were eventually cleared (for example, Fig. 1b, 30 min and Fig. 1c). Most cells contained a juxtanuclear inclusion as well as smaller puncta throughout the cytosol, whereas some cells contained only the juxtanuclear inclusion (Fig. 1b, c). Impairment of proteasome-mediated degradation either in *cim3-1* cells or by treatment with the proteasome inhibitor MG132 stabilized GFP–Ubc9^{ts} and led to its reproducible accumulation in two distinct inclusions in virtually every cell (Fig. 1b, 60 min and 120 min and Supplementary Fig. 1a). At early time points after misfolding in proteasome-defective cells, GFP–Ubc9^{ts} accumulated in structures resembling those observed during degradation in control cells (Fig 1b, compare 15 min and 30 min). Quantification indicated that the juxtanuclear inclusion formed first, closely followed by cytosolic puncta (Fig. 1c). However, at later incubation times at 37 °C the juxtanuclear inclusion remained, but the puncta were no longer observed. Instead, a second large perivacuolar inclusion was now formed at the periphery of the cell (Fig. 1b, c). Once formed, both

[1]Department of Biology and BioX Program, Stanford University, Stanford, California 94305, USA.

inclusions persisted well beyond the time course shown in Fig. 1. Notably, formation of both inclusions was an active process, as it was reversibly inhibited by the microtubule-depolymerizing drug benomyl (Supplementary Fig. 2). These two inclusions may represent distinct compartments for the sequestration of misfolded proteins.

We next examined other types of cytosolic quality control substrates. We initially followed the unassembled von Hippel-Lindau (VHL) tumour suppressor[25,26]. VHL only folds after binding to its cofactor elongin BC[27] (Fig. 1d). Tumour-causing mutations impairing elongin BC binding, or expression in cells lacking elongin BC, lead to misfolded VHL ubiquitination and degradation[25] (Fig. 1d), resulting in reduced levels of diffuse fluorescence (Fig. 1e, compare left panel, misfolded without elongin BC, with right panel, folded VHL with elongin BC). Inhibition of the proteasome in *cim3-1* cells (Fig. 1f), or with MG132 (Supplementary Fig. 1c), led to formation of a single juxtanuclear GFP–VHL inclusion. Importantly, proteasome impairment did not produce GFP–VHL inclusions under conditions leading to productive VHL folding (Fig. 1e, plus elongin BC, right panel).

It was puzzling that at 30 °C VHL consistently formed a single juxtanuclear inclusion whereas Ubc9^{ts} formed two distinct inclusions. Ubc9^{ts} destabilization requires thermal stress, hence formation of two inclusions might result from the increased load of denatured quality control substrates at 37 °C. Indeed, when unassembled VHL was expressed at 37 °C it also accumulated in two inclusions as observed for Ubc9^{ts} (Fig. 1f, c). Three-dimensional fluorescence deconvolution microscopy demonstrated that the inclusions formed by VHL and Ubc9^{ts} overlap spatially in the same compartments (Fig. 1g and Supplementary Movie 1).

A missense mutation of actin, actin(E364K), also degraded via the ubiquitin–proteasome pathway[25], similarly accumulated in the same inclusions as Ubc9^{ts} (Fig. 1g). As clearance of misfolded Ubc9, VHL and actin requires ubiquitination, we considered whether proteasome impairment or stress cause widespread aggregation of ubiquitinated proteins (Supplementary Fig. 1b). This is not the case, as native substrates of the ubiquitin–proteasome pathway[28], such as Arg–GFP (R–GFP), Ub–G76A–GFP (Ub–GFP) and Deg1–GFP (Supplementary Fig. 1b and data not shown), remained soluble and diffuse after proteasome impairment, even under conditions of stress (Supplementary Figs 1b and 3c). We conclude that different classes of misfolded cytosolic proteins are sequestered in two defined cellular inclusions, one juxtanuclear and one at the periphery of the cell. The juxtanuclear inclusion seems to form first and is more prevalent under normal cellular conditions. However, stress conditions lead to protein accumulation in the second peripheral inclusion. In principle, the differential partitioning of non-native quality control substrates between these two compartments may be determined by a change in their intrinsic properties, such as aggregation state, or by their interaction with saturatable quality control components, or both.

We explored the relationship between inclusions formed by disease-related amyloidogenic proteins and those characterized here for

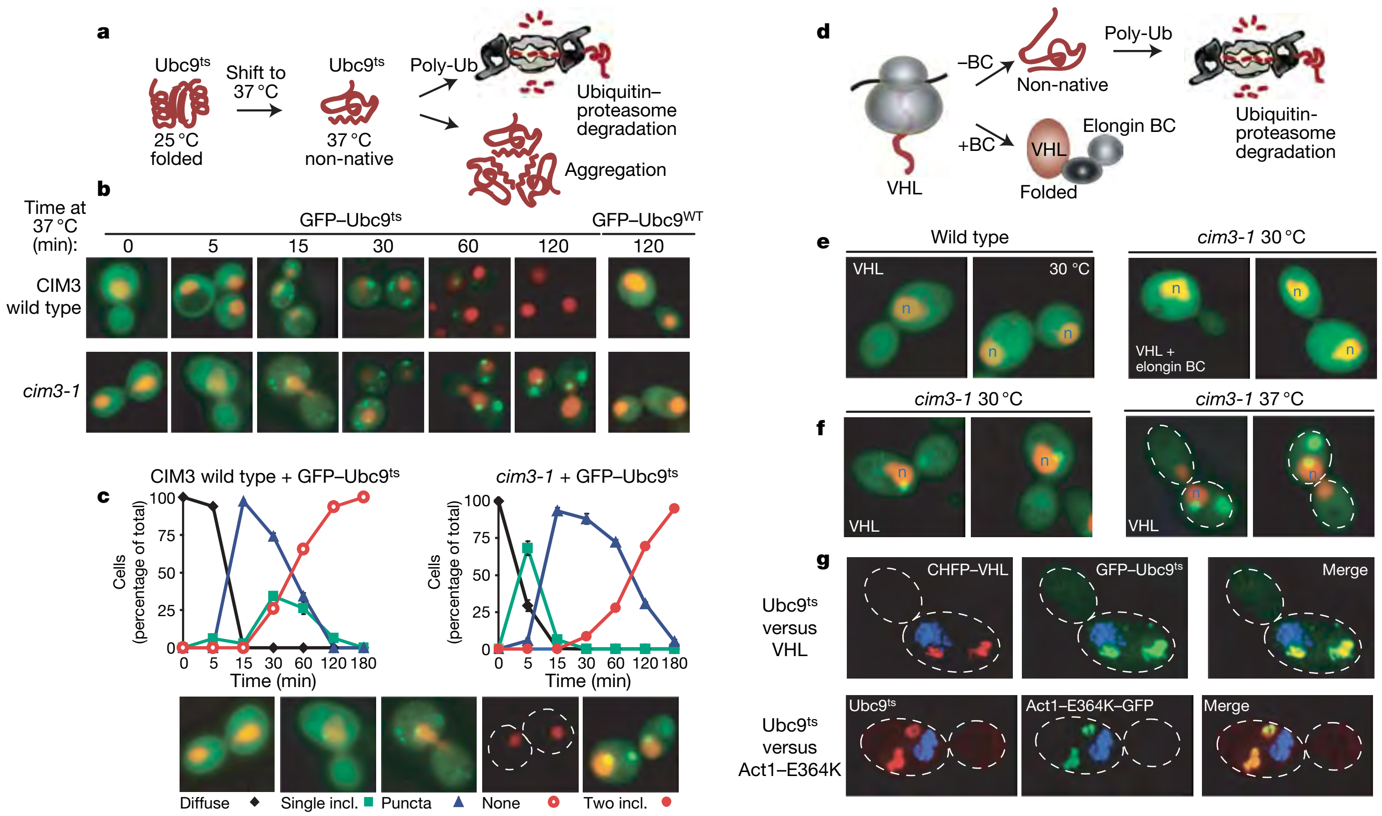


Figure 1 | A panel of quality control substrates defines two distinct compartments for the sequestration of misfolded cytosolic proteins. a, The temperature-sensitive mutant of Ubc9(Y68L) (Ubc9^{ts}) is folded and long-lived at 25 °C. After temperature shift to 37 °C, the Ubc9^{ts} protein misfolds and is degraded by the ubiquitin–proteasome pathway. **b**, Time-dependent changes in localization of folded and misfolded GFP–Ubc9 in wild-type (WT) and *cim3-1* cells. Nuclei were visualized by co-expressing NLS–tdTomato (NLS–TFP). Ubc9 expression was shut off by addition of 2% glucose before temperature shift in all experiments. **c**, Quantification of Ubc9^{ts} localization after misfolding in wild-type and *cim3-1* cells. Graphs represent three separate experiments conducted as in **b**. The phenotypes (see panels) of 100 cells were scored at each time point. **d**, Quality control of the VHL tumour suppressor. VHL folds after elongin BC binding to form the VBC complex. In the absence of elongin BC, VHL is degraded by the ubiquitin–proteasome pathway[25]. **e**, VHL localization in wild-type and *cim3-1* cells, and at 30 °C and 37 °C in *cim3-1* cells (**f**). Two panels are shown for each experiment. **g**, Misfolded VHL, Ubc9 and actin co-localize in the same two inclusions. VHL tagged with mCherry (CHFP–VHL, red) with GFP–Ubc9^{ts} (green, upper panel) or with Act1–E364K–GFP (green, lower panel) in *cim3-1* yeast, after 2 h at 37 °C. Images collected as a Z-series and deconvoluted are shown as a two-dimensional projection.

misfolded cytosolic proteins (Fig. 2a–c). The relative spatial localization of the aggregates formed by glutamine-rich yeast prion proteins Rnq1 and Ure2, as well as polyQ expanded Huntingtin (HttQ103) relative to the Ubc9^{ts} inclusions was determined by deconvolution microscopy. All the amyloidogenic proteins tested formed an inclusion that consistently co-localized with the perivacuolar peripheral inclusion of Ubc9^{ts} (Fig. 2a–c ; Supplementary Movie 2). We did not observe any cases of co-localization of either the prion proteins or Htt with the juxtanuclear inclusion.

Unlike normal quality control substrates, amyloidogenic proteins (including Huntingtin (Htt) and prions) form large insoluble inclusions even in the absence of proteasome inhibition[10,13,17,29]. Thus, amyloidogenic proteins were also analysed in the absence of proteasome inhibition and under normal growth temperatures (Fig. 2d, e). Rnq1, Ure2 and HttQ103 also accumulated under these normal conditions in aggregates localized exclusively in the peripheral compartment (Fig. 2d, e). Additionally, Rnq1 was also found in small puncta throughout the cell (Fig. 2a). The accumulation of amyloidogenic proteins in the peripheral inclusion in the absence of either stress or proteasome impairment (Fig. 2d, e) indicates that this compartment can also form under normal conditions. Notably, Rnq1 always surrounded the Ure2 and HttQ103 deposits (red fluorescence in Fig. 2d, e), suggesting that Rnq1 is targeted to this perivacuolar compartment with slower kinetics than the other amyloidogenic proteins. These observations suggest that some unique feature of amyloidogenic proteins earmarks them for exclusive delivery to the peripheral inclusion.

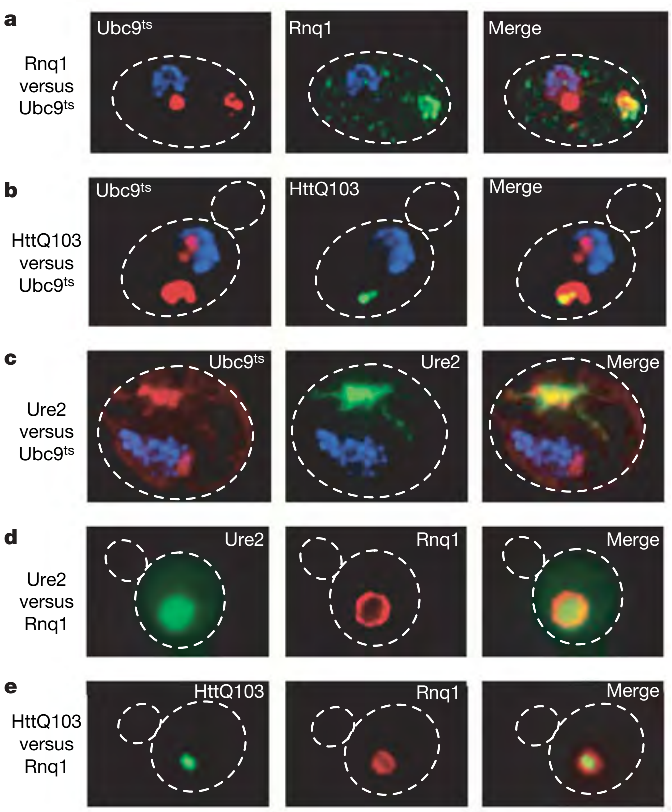


Figure 2 | Amyloidogenic proteins are preferentially directed to a single inclusion. a, Co-localization of inclusions of the yeast prion Rnq1 (green, tagged with GFP) and misfolded Ubc9^{ts} (red, tagged with CHFP), and HttQ103–GFP with CHFP–Ubc9^{ts}(**b**). Ure2–GFP with CHFP–Ubc9^{ts} (**c**), in *cim3-1* yeast after 2 h at 37 °C. Images were collected as a Z-series and deconvoluted. **d**, Co-localization of the yeast prions Ure2–GFP and Rnq1–CHFP in the peripheral inclusion. A direct fluorescence image is shown for **e**. **e**, Co-localization of HttQ103–GFP with Rnq1–CHFP in the peripheral inclusion.

Distinct quality control compartments in Mammalian cells

We next determined whether differential sequestration of misfolded and amyloidogenic proteins in distinct quality control compartments is conserved in mammalian cells (Fig. 3, Supplementary Fig. 1d, e). Misfolded Ubc9^{ts} and VHL showed diffuse fluorescence when expressed in untreated HeLa cells (Fig. 3a, upper panel), with VHL more prominent around the endoplasmic reticulum and nucleus. After proteasome inhibition both proteins co-localized in the perinuclear region, as observed in yeast cells (Fig. 3b, lower panel). Importantly, wild-type folded Ubc9 did not co-aggregate with VHL under these conditions (Supplementary Fig. 1d). We then compared the distribution of HttQ103 and misfolded VHL. As observed in yeast cells, HttQ103 and misfolded VHL were generally sequestered in two different inclusions after proteasome inhibition in mammalian cells (Fig. 3b, lower panel; Supplementary Fig. 1e). In the absence of proteasome impairment, VHL was degraded whereas the HttQ103 inclusion was still observed. It thus appears that the differential sequestration of misfolded proteins in two quality control compartments is conserved from yeast to mammals.

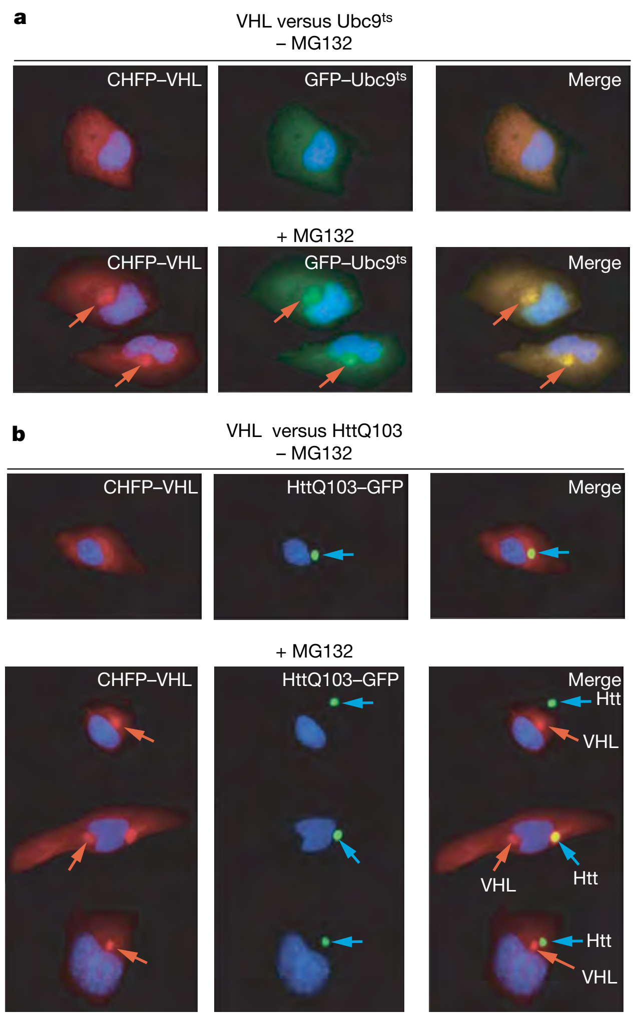


Figure 3 | Mammalian cells differentially sequester misfolded proteins in two distinct compartments. a, CHFP–VHL and GFP–Ubc9^{ts} show diffuse localization in the absence of proteasome inhibition (upper panel), and form co-localizing perinuclear puncta and inclusions next to the endoplasmic reticulum after proteasome inhibition (+MG132, lower panel).
b, HttQ103–GFP forms one hyper-fluorescent inclusion (upper panel, note inclusion not always perinuclear). After proteasome inhibition (+MG132), CHFP–VHL forms an inclusion that is distinct from that of HttQ103–GFP (lower panel).

Quality control compartments show distinct diffusion properties

We next examined the solubility state of misfolded proteins in either inclusion by determining their diffusion properties using Fluorescence Loss in Photobleaching (FLIP)[30]. In brief, a laser pulse was used to photobleach GFP–Ubc9^{ts} from a small section of cytosol outside of the two GFP–Ubc9^{ts} inclusions (Fig. 4a, square). The ensuing changes in fluorescence intensity of the different cellular compartments, assessed as a function of time, provide a measure of their relative exchange rate with the bleached cytoplasmic portion (Fig. 4a). Bleaching caused a rapid loss of the diffuse cytosolic fluorescence corresponding to soluble GFP–Ubc9^{ts} (Fig. 4a, black trace). A rapid fluorescence loss was also observed for the juxtanuclear inclusion (Fig. 4a, red trace), indicating that a substantial fraction of GFP–Ubc9^{ts} in this compartment is soluble and can exchange with the cytosolic pool. We therefore refer to this inclusion as the 'juxtanuclear quality control' compartment, or JUNQ. In contrast, following a small initial reduction in fluorescence, the peripheral perivacuolar compartment retained most ($>70\%$) of its fluorescent signal (Fig. 4a, blue trace). This suggests that this inclusion contains a large fraction of non-diffusing, possibly insoluble GFP–Ubc9^{ts}. Accordingly, we named this inclusion the 'insoluble protein deposit', or IPOD. As the conservation of fluorescent signal within the IPOD could result from a barrier to exchange with the cytosolic pool, for example, by a membrane, we examined the internal mobility of the protein within the IPOD using FRAP[30]. When a small sector within the IPOD was directly bleached, we did not observe any redistribution of the fluorescent signal within the IPOD from the non-bleached part of the inclusion. This indicates that the protein in this structure is immobile, consistent with this compartment containing aggregated species (Fig. 4b).

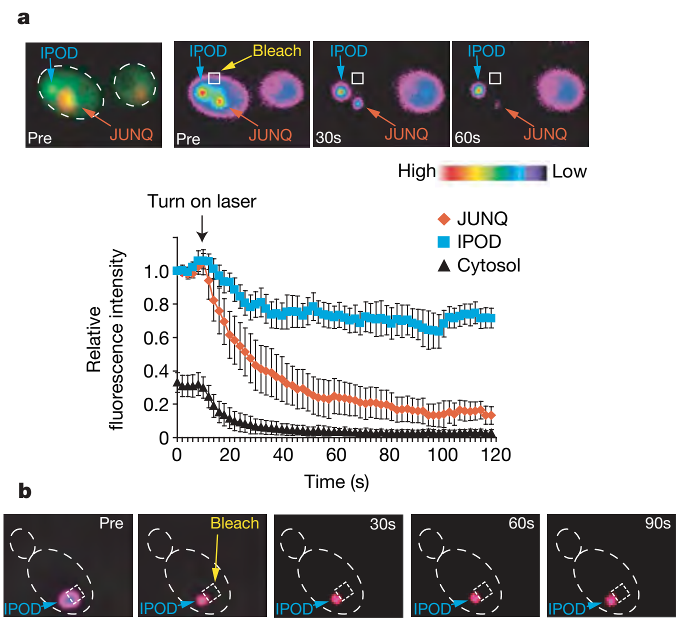


Figure 4 | Differential solubility of misfolded substrates in the distinct quality control compartments. a, Qualitative FLIP analysis indicates that misfolded protein in the JUNQ and the IPOD exhibit different relative exchange rates with the cytosolic pool. Pre- and post-bleach images of a representative FLIP experiment with GFP–Ubc9^{ts} are shown. The fluorescence intensity scale is pseudocoloured as shown. A square designates the location of the photobleaching laser spot. GFP–Ubc9^{ts} was expressed in *cim3-1* yeast. Relative fluorescence of JUNQ (orange), IPOD (blue) and cytosol (black) from ten FLIP experiments is shown over time. **b**, Protein in the IPOD inclusion is immobile. Pre- and post-bleach images of a representative FRAP experiment and subsequent recovery of GFP–Ubc9^{ts} are shown. GFP–Ubc9^{ts} was expressed in *Δubc4/5* cells and shifted to 37 °C to form the IPOD.

Biochemical analyses supported the conclusion that the JUNQ and IPOD accumulate proteins in distinct solubility states (Supplementary Fig. 3). VHL localized only to the JUNQ was in a Triton-soluble state (Supplementary Fig. 3a), whereas accumulation in the IPOD correlated with a shift to the insoluble fraction (Supplementary Fig. 3a; see also Fig. 6b and Supplementary Fig. 6g for Ubc9 and Htt). We conclude that one compartment, the JUNQ, contains a large fraction of soluble misfolded protein, whereas the IPOD compartment contains non-diffusing, insoluble species. The observation that amyloidogenic proteins appear to be targeted exclusively to the IPOD suggests that this compartment is the preferred cellular destination for protein aggregates.

JUNQ and IPOD are defined subcellular compartments

Because different quality control substrates reproducibly accumulated in the same two compartments, we examined the relationship of both JUNQ and IPOD with known cellular structures and components (Fig. 5 and Supplementary Fig. 5). Deconvolution microscopy indicated that the JUNQ is formed in an indentation of the nucleus (Fig. 5a; see also Supplementary Movies 1 and 2). In dividing cells both JUNQ and IPOD were invariably retained in the mother cell, raising the possibility that these compartments provide a mechanism to retain misfolded proteins in the mother cells during cell division (Supplementary Fig. 4a and data not shown). This could explain previous observations indicating that oxidatively damaged proteins are prevented from entering the daughter cell[31,32]. Notably, neither the JUNQ nor the IPOD were localized to the spindle pole body (Fig. 5b) unlike the aggresome which co-localizes with the microtubule organizing centre[20].

A similar analysis using the endoplasmic reticulum marker Sec63 (ref. 19) indicated that the JUNQ is in close proximity to the endoplasmic reticulum. Sec63 redistributed around the JUNQ relative to the remaining nuclear envelope (Fig. 5c; Supplementary Fig. 4). Sec63 usually marks the ERAC structures that accumulate ERAD substrates[19], suggesting that the JUNQ forms at a defined cellular location in close proximity to the region that participates in ERAD. Perhaps the localization of cytosolic and endoplasmic reticulum misfolded proteins to one cellular location serves to concentrate cellular quality control components with their substrates to both enhance the efficiency of misfolded protein clearance and sequester them from the cellular milieu.

Given the centrality of proteasomal degradation in protein turnover, we examined the cellular distribution of 26S proteasomes using previously characterized GFP-tagged proteasomes (Fig. 5d and Supplementary Fig. 4c–e)[33]. Deconvolution microscopy revealed that most proteasomes in the cells localize to the endoplasmic reticulum surface, even though there is a fraction of diffuse proteasomes. For all misfolded proteins examined, we observed a re-distribution of proteasomes to the site of JUNQ protein accumulation both at 30 °C and 37 °C (Fig. 5d; Supplementary Fig. 4c–e). Mutant actin occasionally accumulated in the quality control compartments even without proteasome inhibition (Fig. 5d, bottom). This also led to proteasome re-distribution to the JUNQ, suggesting that misfolded protein accumulation can recruit proteasomes to the JUNQ. In contrast, proteasomes did not co-localize with the IPOD, indicating that soluble misfolded proteins, rather than insoluble amyloid aggregates, cause a re-distribution of cellular proteasomes. Our data suggest that the perinuclear JUNQ compartment acts as a major site of proteasome concentration and misfolded protein degradation.

The chaperone Hsp104, another quality control component that interacts with misfolded and aggregated proteins, was examined using a functional GFP-tagged Hsp104 (ref. 34). Hsp104 co-localized with both JUNQ and IPOD (Fig. 5e; Supplementary Fig. 4e–g). Most Hsp104 accumulated around the IPOD compartment, often in an arrangement around the protein inclusion (Fig. 5e). Hsp104 was often found in IPOD- or JUNQ- like inclusions in the absence of ectopically expressed misfolded proteins (Fig. 5e, middle panel, and data not shown), indicating that these quality control compartments are

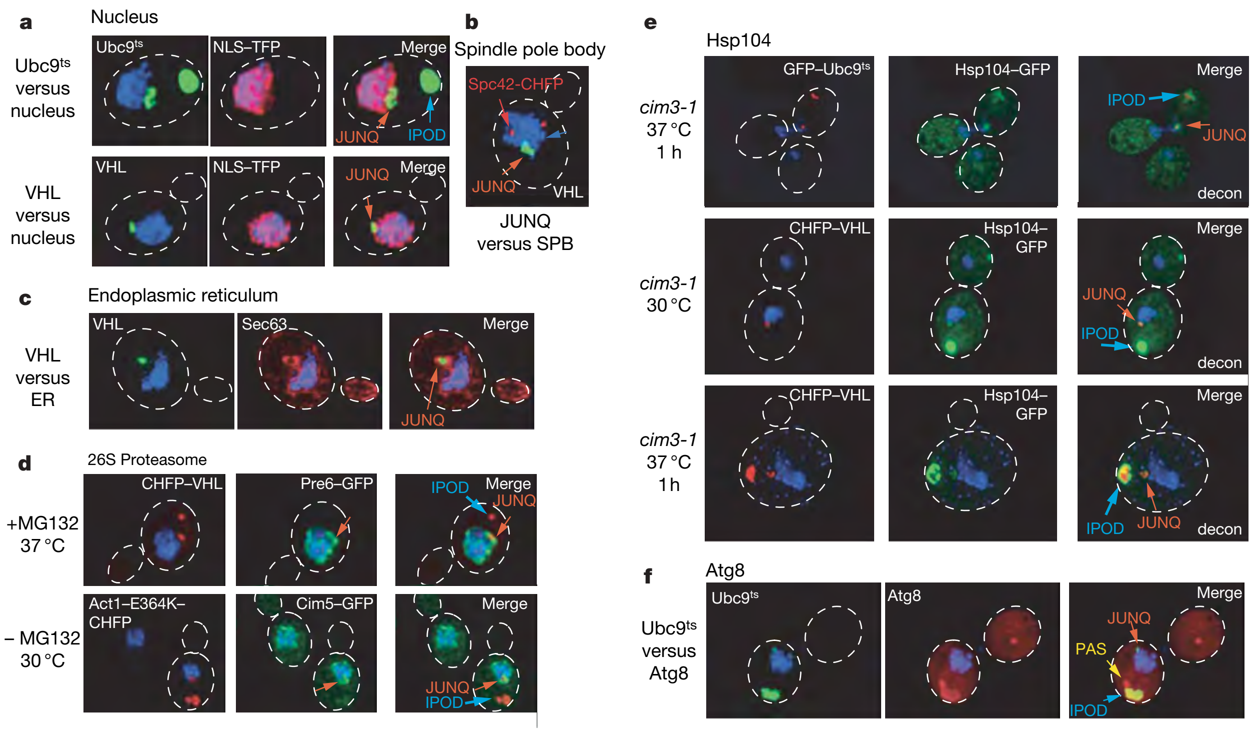


Figure 5 | The JUNQ and IPOD are defined subcellular compartments. **a**, JUNQ compartment tightly co-localizes with nuclear membrane. DNA is visualized with DAPI (blue), nucleoplasm with NLS–TFP (red). GFP–Ubc9^{ts} (top) at 37 °C and GFP–VHL (bottom) in *cim3-1* cells. Two-dimensional projections of de-convoluted Z-series are shown. **b**, The JUNQ compartment, shown here for GFP–VHL, does not localize to the spindle pole body (marked by Spc42–CHFP), and is in close proximity to the endoplasmic reticulum visualized with Sec63–CHFP (**c**). **d**, The JUNQ compartment (upper panel, CHFP–VHL, red), but not the IPOD (Act1–E364K–CHFP, lower panel), concentrates 26S proteasomes (visualized with Cim5–GFP for regulatory particle and Pre6–GFP for core particle, green). **e**, Hsp104 localizes to both compartments. JUNQ and IPOD were formed by expressing CHFP–Ubc9^{ts} (upper panel) or CHFP–VHL (lower panel). Note Hsp104 also accumulates in an IPOD structure (blue arrow) independently of ectopically-expressed aggregating protein when CHFP–VHL is expressed in *cim3-1* cells at 30 °C (middle panel). **f**, The IPOD, shown here for GFP–Ubc9^{ts}, co-localizes with CHFP–Atg8. Some CHFP–Atg8 can also be seen in the pre-autophagosomal structure (PAS)[12].

normally present in cells. The co-localization of Hsp104 with the IPOD resonates with its role in disaggregating or fragmenting aggregates of prion proteins[4], which also localize to this compartment, and with its role in preventing inheritance of oxidatively damaged proteins[31,32]. At the JUNQ, Hsp104 may serve to keep proteins soluble for either refolding or degradation (Fig. 6e).

Autophagy has been implicated in the clearance of protein aggregates[35] and possibly also ubiquitinated misfolded proteins[36–38]. The IPOD, but not the JUNQ, co-localized with the autophagic marker Atg8 (ref. 12) and the pre-autophagosomal structure (PAS; Fig. 5f; see also Supplementary Fig. 5 for electron microscopy). The association of the IPOD with an autophagic marker (Supplementary Fig. 5b for electron microscopy) provides an intriguing link between aggregated proteins in the IPOD and the autophagy pathway. Neither Atg8 nor Atg7, both essential components of the autophagic pathway, were required for IPOD formation (Supplementary Fig. 5c). Thus, it is also possible that that Atg8 co-localization has an autophagy-independent function in the IPOD, as recently proposed for its mammalian homologue LC3 (refs 36–38).

Immunoelectron microscopy analysis confirmed that the JUNQ is closely associated with the nucleus, and may be flanked by proliferations of the nuclear membrane (Supplementary Fig. 5a, b). The IPOD (Supplementary Fig. 5a) was made up of electron-dense material, consistent with our FRAP and biochemical characterization. Occasionally, both electron microscopy and fluorescence analyses of prion IPODs showed labelling in a circular hollow pattern around a densely packed core (Supplementary Fig. 5a, Ure2–GFP). We hypothesize that the ectopically expressed prions are occasionally layered over aggregates in pre-existing IPODs containing endogenous proteins (for example, Hsp104 in Fig. 5e).

Role of ubiquitination in substrate partitioning

We considered whether ubiquitination of misfolded proteins[15] plays a role in their partitioning to either the JUNQ or the IPOD (Fig. 6 and Supplementary Fig. 6a). Degradation of misfolded VHL, Ubc9 and actin requires ubiquitination by the E2 pair Ubc4/5 (refs 23, 25 and 26). Ubiquitination was impaired by expressing these proteins in *Δubc4/5* cells (Fig. 6a, VHL; Supplementary Fig. 6,Ubc9^{ts}; actin, data not shown) or by overexpression of Ubp4 (ref. 39; Supplementary Fig. 6e). Similar results were obtained using either strategy for all misfolded proteins (Fig. 6a; Supplementary Fig. 6b, c). Impairing misfolded protein ubiquitination blocked their accumulation in the JUNQ and instead resulted in exclusive accumulation in the IPOD, even at 30 °C and in the absence of proteasome inhibition (Fig. 6, Supplementary Fig. 6). The IPODs formed under these conditions exhibited the same morphology as observed previously, and were also Atg8 and Hsp104-positive (Supplementary Figs 6d and 7e).

Consistent with its relocalization to the IPOD, blocking ubiquitination reduced the solubility of misfolded VHL (Fig. 6b) and Ubc9 (Supplementary Fig. 6g). Misfolded VHL was Triton-soluble under conditions where it only forms the JUNQ (Fig. 6c, left panel), but was almost entirely insoluble when targeted to the IPOD once ubiquitination was blocked (Fig. 6b, right panel). Similar results were obtained with Ubc9^{ts}; blocking its ubiquitination rendered it as insoluble as aggregated HttQ53 (Supplementary Fig. 6g). Alternatively, blocking the ubiquitination of a native degradation substrate, Ub–GFP, in the same *Δubc4/5* cells did not impair its solubility (Supplementary Fig. 3c). Thus, ubiquitination is an important determinant for maintaining solubility of misfolded proteins and sorting them to the JUNQ, whereas non-ubiquitinated species are directed to the IPOD.

We next exploited the observation that the chaperone Sti1 is required for VHL degradation, but not for VHL folding[25]. Sti1 deletion also directed VHL to the IPOD, as observed in *Δubc4/5* cells (Supplementary Fig. 7d, e), but had no effect on the localization of Ubc9^{ts} (Supplementary Fig. 7c). This result indicates that ubiquitination and partitioning of misfolded proteins between JUNQ and IPOD is modulated by specific interactions with the cellular chaperone network. The finding that amyloidogenic proteins are primarily targeted to the IPOD may thus reflect their inefficient interaction with quality control chaperone and ubiquitination components. This property could distinguish prions and other amyloidogenic proteins from the bulk of misfolded quality control substrates that normally do not accumulate in amyloids.

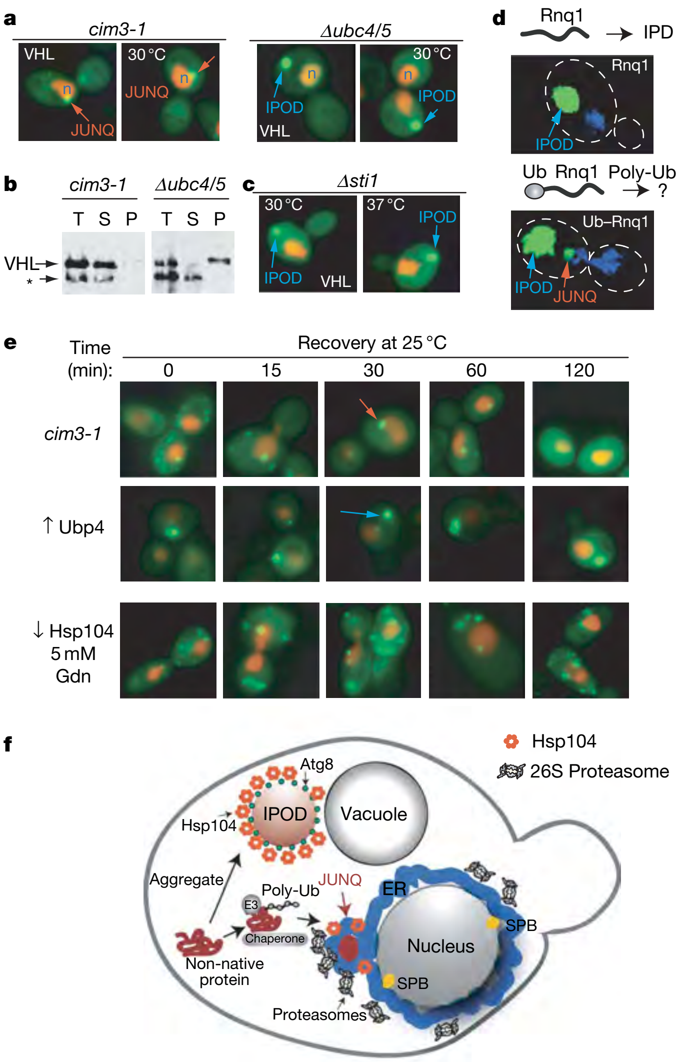


Figure 6 | Partitioning between JUNQ and IPOD is regulated by ubiquitination. a, Blocking ubiquitination of misfolded VHL prevents its localization to JUNQ, and redirects these proteins to the IPOD. **b**, VHL in the IPOD accumulates in a Triton-insoluble fraction. Asterisk denotes cross-reacting band unrelated to VHL. **c**, Deletion of Sti1, required for VHL degradation[25], reroutes misfolded VHL to the IPOD. **d**, Ubiquitination suffices to promote prion delivery to the JUNQ. Rnq1–GFP localizes exclusively to the IPOD. A ubiquitination signal (Ub–G76A) engineered in the yeast prion Rnq1–GFP causes Ub–G76A–Rnq1–GFP localization to both JUNQ and IPOD. **e**, Recovery of diffuse cytosolic fluorescence by thermally denatured GFP–Ubc9^{ts} accumulated in the JUNQ (top) but not in the IPOD (middle) after return to the permissive temperature. The recovery of GFP–Ubc9^{ts} requires Hsp104 disaggregase activity (bottom). **f**, Model for sorting of cytosolic misfolded proteins to distinct quality control compartments.

Unexpectedly, blocking the ubiquitination of misfolded proteins caused them to behave like amyloidogenic proteins. We considered whether enhancing ubiquitination of the prion Rnq1 by engineering a synthetic ubiquitination signal suffices to promote partitioning to the JUNQ (Fig. 6d, Supplementary Fig. 8). Rnq1 normally only accumulates in IPOD inclusions (Figs 2b–f and 6d, left panel), but enhancing its ubiquitination with an amino-terminal ubiquitin-fusion degradation (UFD) signal (Supplementary Fig. 8a) directed a fraction of Ub–Rnq1 to the JUNQ (Fig. 6d, right panel). Biochemical analysis confirmed this observation (Supplementary Fig. 8b). Rnq1 is predominantly in the insoluble fraction (Supplementary Fig. 8b, left panel) whereas Ub–Rnq1 is distributed between soluble and insoluble fractions (Supplementary Fig. 8b, right panel). Notably, higher molecular weight poly-ubiquitinated Ub–Rnq1 was found only in the soluble fraction. As non-ubiquitinated Ub–Rnq1 nevertheless has one N-terminal ubiquitin, it seems that poly-ubiquitination is required for Rnq1 delivery to the JUNQ. These experiments indicate that poly-ubiquitination is a key determinant for partitioning misfolded proteins between the two quality control compartments, as it is necessary for sorting misfolded proteins to the JUNQ and sufficient to redirect a prion protein from the IPOD to the JUNQ.

Consequences of protein accumulation in the JUNQ or IPOD

We next examined the functional consequences of accumulation in the JUNQ or the IPOD (Fig. 6e) by exploiting the observation that the thermal denaturation of Ubc9^{ts} is reversible (data not shown)[24]. Thermally denatured GFP–Ubc9^{ts} was directed to either the JUNQ or the IPOD by changing its ubiquitination state; the fate of Ubc9^{ts} in either compartment was then examined after return to the permissive temperature (Fig. 6e). Cells containing Ubc9^{ts} in puncta and the JUNQ gradually recovered diffuse GFP–Ubc9^{ts} fluorescence (Fig. 6e, top), In contrast, cells with GFP–Ubc9^{ts} in the IPOD did not recover diffuse fluorescence (Fig. 6e, data not shown for *Δubc4/5*). Ubc9^{ts} refolding from the JUNQ required Hsp104 and was inhibited by 5 mM guanidine, suggesting that Hsp104 facilitates refolding of proteins in the JUNQ. Thus, misfolded Ubc9^{ts} that is ubiquitinated and sorted into the JUNQ can be refolded by the cellular chaperone machinery, whereas Ubc9^{ts} sorted to the IPOD is terminally sequestered from the cytoplasm. These experiments suggest that poly-ubiquitination not only targets proteins for degradation, but may also contribute to their re-folding competence.

Discussion

Cellular quality control networks have a key role in maintaining protein homeostasis[40]. We find two cytosolic quality control compartments, the JUNQ and the IPOD, carry out general but distinct functions managing misfolded and aggregated proteins and are conserved from yeast to mammalian cells (Fig. 6f). After misfolding, most proteins will be recognized and ubiquitinated by the quality control machinery, which directs them to the JUNQ, a region that concentrates disaggregating chaperones and 26S proteasomes and is in close proximity to the perinuclear endoplasmic reticulum region involved in ERAD (Fig. 6f). Accumulation of misfolded proteins and quality control components at the endoplasmic reticulum membrane may facilitate both degradation and refolding by increasing their local concentrations and enhancing encounter rates by restricting diffusion. Cellular increase of misfolded protein loads, for example, by stress or during ageing, may saturate the quality control machinery needed for sorting to the JUNQ, resulting in accumulation of aggregated and potentially toxic species. These misfolded proteins are directed to the IPOD, which seems to terminally sequester protein aggregates. The spatial sequestration of these aggregates from the site where most proteasomal degradation takes place may serve a protective function. Spatial sequestration may also facilitate aggregate clearance, either through the autophagic pathway or by dilution through retention in the dividing mother cell.

The identification of two distinct quality control compartments resonates with a number of previous observations. Studies in *Caenorhabditis elegans* suggest two hierarchical pathways for degradation of amyloidogenic Aβ species[41,42]. Alternative degradation pathways were reported for ERAD substrates[21] and α-synuclein[43], with soluble species of the same protein degraded via the proteasome and insoluble aggregates by autophagy[43,44]. For mammalian cells, there have been disparate reports on the solubility and structural properties of protein inclusions[13,16,18,36]. Thus, polyQ inclusions in mammalian cells are immobile[17], similar to the IPOD in yeast. The distinct localization and solubility of misfolded proteins in the JUNQ also resembles observations that mammalian SOD1 inclusions contain soluble protein and are spatially distinct from polyQ aggregates[13,16]. Soluble misfolded VHL is also reported to accumulate in an ubiquitination-dependent manner on the cytosolic surface of the mammalian endoplasmic reticulum, similar to our findings in yeast[13,16]. Importantly, our findings provide a framework for integrating these various observations into a conserved cellular pathway of quality control. Further, we identify solubility and the ubiquitination state of a quality control substrate as key determinants of its delivery to either the JUNQ or the IPOD. Future studies should determine whether poly-ubiquitination exerts these effects through interactions with cellular components or by directly enhancing protein solubility.

The preferential targeting of amyloidogenic proteins to the IPOD may arise from their reduced affinity for quality control components. In turn, this may lead to higher levels of toxic misfolded conformations, consistent with findings that the toxicity of amyloidogenic proteins resides in small soluble species[45]. Accordingly, overexpression of chaperones and ubiquitination components, which alleviate toxicity, could compensate for their reduced affinity for the amyloidogenic species[4,22,46,47]. The observation that amyloidogenic and globular misfolded proteins are differentially engaged by cellular quality control pathways provides new perspectives on the molecular basis of protein conformation diseases, and may have useful implications for the understanding of neurodegeneration, ageing and stress.

METHODS SUMMARY

Yeast growth, manipulation and protein expression were performed as described[25,48,49]. Misfolded and amyloidogenic proteins used in this study were tagged with EGFP, mCherry or tdTomato[50]. Conventional epifluorescence micrographs were obtained from live yeast cells on a Zeiss Axiovert microscope with a ×100 oil lens (NA1.4). Deconvoluted images were acquired using an Olympus microscope. Digital images (12 bit) were digitally deconvoluted by using DELTAVISION software (Applied Precision). Live-cell imaging was performed using the Marianas system from Intelligent Imaging Innovations equipped with the MicroPoint FRAP laser system (Photonic Instruments).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 January 2007; accepted 24 June 2008.

1. Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292**, 1552–1555 (2001).
2. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).
3. Gidalevitz, T., Ben-Zvi, A., Ho, K. H., Brignull, H. R. & Morimoto, R. I. Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* **311**, 1471–1474 (2006).
4. Muchowski, P. J. & Wacker, J. L. Modulation of neurodegeneration by molecular chaperones. *Nature Rev. Neurosci.* **6**, 11–22 (2005).
5. Outeiro, T. F. & Lindquist, S. Yeast cells provide insight into α-synuclein biology and pathobiology. *Science* **302**, 1772–1775 (2003).
6. Schaffar, G. *et al.* Cellular toxicity of polyglutamine expansion proteins: mechanism of transcription factor deactivation. *Mol. Cell* **15**, 95–105 (2004).
7. Lesne, S. *et al.* A specific amyloid-β protein assembly in the brain impairs memory. *Nature* **440**, 352–357 (2006).
8. Rubinsztein, D. C. The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature* **443**, 780–786 (2006).
9. Iwata, A. *et al.* Increased susceptibility of cytoplasmic over nuclear polyglutamine aggregates to autophagic degradation. *Proc. Natl Acad. Sci. USA* **102**, 13135–13140 (2005).
10. Taylor, J. P. *et al.* Aggresomes protect cells by enhancing the degradation of toxic polyglutamine-containing protein. *Hum. Mol. Genet.* **12**, 749–757 (2003).
11. Rideout, H. J., Lang-Rollin, I. & Stefanis, L. Involvement of macroautophagy in the dissolution of neuronal inclusions. *Int. J. Biochem. Cell Biol.* **36**, 2551–2562 (2004).
12. Yorimitsu, T. & Klionsky, D. J. Autophagy: molecular machinery for self-eating. *Cell Death Differ.* **12** (Suppl 2), 1542–1552 (2005).
13. Matsumoto, G., Kim, S. & Morimoto, R. I. Huntingtin and mutant SOD1 form aggregate structures with distinct molecular properties in human cells. *J. Biol. Chem.* **281**, 4477–4485 (2006).
14. Bucciantini, M. *et al.* Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507–511 (2002).
15. Sherman, M. Y. & Goldberg, A. L. Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. *Neuron* **29**, 15–32 (2001).
16. Matsumoto, G., Stojanovic, A., Holmberg, C. I., Kim, S. & Morimoto, R. I. Structural properties and neuronal toxicity of amyotrophic lateral sclerosis-associated Cu/Zn superoxide dismutase 1 aggregates. *J. Cell Biol.* **171**, 75–85 (2005).
17. Krobitsch, S. & Lindquist, S. Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins. *Proc. Natl Acad. Sci. USA* **97**, 1589–1594 (2000).
18. Kamhi-Nesher, S. *et al.* A novel quality control compartment derived from the endoplasmic reticulum. *Mol. Biol. Cell* **12**, 1711–1723 (2001).
19. Huyer, G. *et al.* A striking quality control subcompartment in *Saccharomyces cerevisiae*: the endoplasmic reticulum-associated compartment. *Mol. Biol. Cell* **15**, 908–921 (2004).
20. Kopito, R. R. Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* **10**, 524–530 (2000).
21. Kruse, K. B., Brodsky, J. L. & McCracken, A. A. Characterization of an *ERAD* gene as *VPS30/ATG6* reveals two alternative and functionally distinct protein quality control pathways: one for soluble Z variant of human α-1 proteinase inhibitor (A1PiZ) and another for aggregates of A1PiZ. *Mol. Biol. Cell* **17**, 203–212 (2006).
22. McClellan, A. J., Tam, S., Kaganovich, D. & Frydman, J. Protein quality control: chaperones culling corrupt conformations. *Nature Cell Biol.* **7**, 736–741 (2005).
23. Betting, J. & Seufert, W. A yeast Ubc9 mutant protein with temperature-sensitive *in vivo* function is subject to conditional proteolysis by a ubiquitin- and proteasome-dependent pathway. *J. Biol. Chem.* **271**, 25790–25796 (1996).
24. Tongaonkar, P., Beck, K., Shinde, U. P. & Madura, K. Characterization of a temperature-sensitive mutant of a ubiquitin-conjugating enzyme and its use as a heat-inducible degradation signal. *Anal. Biochem.* **272**, 263–269 (1999).
25. McClellan, A. J., Scott, M. D. & Frydman, J. Folding and quality control of the VHL tumor suppressor proceed through distinct chaperone pathways. *Cell* **121**, 739–748 (2005).
26. Vang, S. *et al.* Actin mutations in hypertrophic and dilated cardiomyopathy cause inefficient protein folding and perturbed filament formation. *FEBS J.* **272**, 2037–2049 (2005).
27. Feldman, D. E., Thulasiraman, V., Ferreyra, R. G. & Frydman, J. Formation of the VHL-elongin BC tumor suppressor complex is mediated by the chaperonin TRiC. *Mol. Cell* **4**, 1051–1061 (1999).
28. Mateus, C. & Avery, S. V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* **16**, 1313–1323 (2000).
29. Duennwald, M. L., Jagadish, S., Giorgini, F., Muchowski, P. J. & Lindquist, S. A network of protein interactions determines polyglutamine toxicity. *Proc. Natl Acad. Sci. USA* **103**, 11051–11056 (2006).
30. Lippincott-Schwartz, J. & Patterson, G. H. Development and use of fluorescent protein markers in living cells. *Science* **300**, 87–91 (2003).
31. Rujano, M. A. *et al.* Polarised asymmetric inheritance of accumulated protein damage in higher eukaryotes. *PLoS Biol.* **4**, e417 (2006).
32. Erjavec, N. & Nystrom, T. Sir2p-dependent protein segregation gives rise to a superior reactive oxygen species management in the progeny of *Saccharomyces cerevisiae. Proc. Natl Acad. Sci. USA* **104**, 10877–10881 (2007).
33. Enenkel, C., Lehmann, A. & Kloetzel, P. M. Subcellular distribution of proteasomes implicates a major location of protein degradation in the nuclear envelope-ER network in yeast. *EMBO J.* **17**, 6144–6154 (1998).
34. Tkach, J. M. & Glover, J. R. Amino acid substitutions in the C-terminal AAA+ module of Hsp104 prevent substrate recognition by disrupting oligomerization and cause high temperature inactivation. *J. Biol. Chem.* **279**, 35692–35701 (2004).
35. Sarkar, S. *et al.* Small molecules enhance autophagy and reduce toxicity in Huntington's disease models. *Nature Chem. Biol.* **3**, 331–338 (2007).
36. Szeto, J. *et al.* ALIS are stress-induced protein storage compartments for substrates of the proteasome and autophagy. *Autophagy* **2**, 189–199 (2006).
37. Kuma, A., Matsui, M. & Mizushima, N. LC3, an autophagosome marker, can be incorporated into protein aggregates independent of autophagy: caution in the interpretation of LC3 localization. *Autophagy* **3**, 323–328 (2007).
38. Pankiv, S. *et al.* p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *J. Biol. Chem.* **282**, 24131–24145 (2007).
39. Swaminathan, S., Amerik, A. Y. & Hochstrasser, M. The Doa4 deubiquitinating enzyme is required for ubiquitin homeostasis in yeast. *Mol. Biol. Cell* **10**, 2583–2594 (1999).
40. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science* **319**, 916–919 (2008).

41. Siegel, S. J., Bieschke, J., Powers, E. T. & Kelly, J. W. The oxidative stress metabolite 4-hydroxynonenal promotes Alzheimer protofibril formation. *Biochemistry* **46,** 1503–1510 (2007).
42. Cohen, E., Bieschke, J., Perciavalle, R. M., Kelly, J. W. & Dillin, A. Opposing activities protect against age-onset proteotoxicity. *Science* **313,** 1604–1610 (2006).
43. Webb, J. L., Ravikumar, B., Atkins, J., Skepper, J. N. & Rubinsztein, D. C. α-Synuclein is degraded by both autophagy and the proteasome. *J. Biol. Chem.* **278,** 25009–25013 (2003).
44. Kruse, K. B., Brodsky, J. L. & McCracken, A. A. Autophagy: an ER protein quality control process. *Autophagy* **2,** 135–137 (2006).
45. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431,** 805–810 (2004).
46. Barral, J. M., Broadley, S. A., Schaffar, G. & Hartl, F. U. Roles of molecular chaperones in protein misfolding diseases. *Semin. Cell Dev. Biol.* **15,** 17–29 (2004).
47. Tam, S., Geller, R., Spiess, C. & Frydman, J. The chaperonin TRiC controls polyglutamine aggregation and toxicity through subunit-specific interactions. *Nature Cell Biol.* **8,** 1155–1162 (2006).
48. Melville, M. W., McClellan, A. J., Meyer, A. S., Darveau, A. & Frydman, J. The Hsp70 and TRiC/CCT chaperone systems cooperate *in vivo* to assemble the von Hippel-Lindau tumor suppressor complex. *Mol. Cell. Biol.* **23,** 3141–3151 (2003).
49. Adams, A., Gottschling, D., Kaiser, C. & Stearns, T. *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory Press, 1997).
50. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnol.* **22,** 1567–1572 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Tsien, S. Michaelis, J. Glover, C. Enenkel and V. Albanese for reagents; J. Mulholland for electron microscopy and deconvolution microscopy help; S. Yamada and W. J. Nelson for help with the live cell microscopy. We thank R. Andino, W. Burkholder, J. England, R. Geller, M. Kaganovich, E. Bennett, J. Nelson and members of the Frydman laboratory for discussions and comments on the manuscript. This work was supported by grants from NIH to R.K. and J.F.

Author Contributions J.F. directed the project; D.K. and J.F. designed and interpreted all experiments; R.K. participated in the initial fluorescence microscopy experiments; D.K. performed all experiments. D.K. and J.F. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.F. (jfrydman@stanford.edu).

doi:10.1038/nature07195

METHODS

Yeast media, plasmids, and strains. Yeast media preparation, growth, transformations and manipulations were performed according to standard protocols[49]. The protein substrates used in this study were visualized as fusions to fluorescent proteins derived from GFP. GFP–Ubc9[ts], GFP–Ubc9[WT], GFP–VHL, Act1–E364K–GFP, Rnq1–GFP, Ure2–GFP, Ub–G76A–GFP, Ub–Arg–GFP, Ub–G76A–Rnq1–GFP, CHFP–Apg8, NLS–tdTomato[50] (TFP) were cloned into pESC (GAL1 URA3; Stratagene). Each of the above was also cloned into pESC GAL1 LEU2 vectors, and identical fusion proteins were made with mCherry[50] fluorescent protein (CHFP) instead of GFP. The pESC plasmid expressing elongin B and elongin C from a GAL-inducible promoter is described elsewhere[25,48]. All proteins were cloned by PCR from yeast genomic DNA or a template plasmid and verified by sequencing. Spc42–GFP and Spc42–CHFP were cloned downstream of the Tub2 promoter into the pRS316 vector. Sec63–CHFP was cloned by excising the GFP from pSM1462 (ref. 51) and replacing it with mCherry. GFP–Hsp104 (ref. 34) was a gift from J. Glover.

The yeast strains used in this study are as follows: WT CIM3 (YPG499; MAT**a** *ura3-52 leu2-Δ1 his3-Δ200 trp1-Δ63 lys2-801 ade2-101*) and *cim3-1* (CMY762; *ura3-52 leu2-Δ1 his3-Δ200 cim3-1*)[52]; MHY501 (MATα *his3-Δ200 leu2-3, 112 ura3-52 lys2-801 trp1-1*) and the isogenic mutant strains MHY508 *(ubc4::HIS3 ubc5::LEU2)* and MHY570 (*ubc4::TRP1 ubc5::LEU2 ubc6::HIS3 ubc7::LEU2*)[53]; GCE6 (*MAT***a** *his3-11,15 leu2-3,112 ura3 PRE6-GFPHA::HIS3::URA3*), GAL5 (*MAT***a** *his3-11,15 leu2-3,112 ura3 CIM5-GFPHA::HIS3::URA3*)[33]; YKO WT, *Δsti1*, *Δpdr5* (MAT**a**/MATα *orfΔ::kanMX4/orfΔ::kanMX4 ura3Δ0/ura3Δ0 leu2Δ0/leu2Δ0 his3Δ1/his3Δ1 met15Δ0/MET15 lys2Δ0/LYS2* (*Saccharomyces* Genome Project))[54]. For experiments using MG132, the YKO (B4147) strains lacking the Pdr5 transporter were used as wild type. Deletion of Pdr5 sensitizes cells to the proteasome inhibitor MG132. When indicated, cells were treated with 80 μM MG132 (Sigma) dissolved in DMSO for 1 h. For all experiments, expression was shut off before temperature shift and microscopy by addition of 2 % glucose.

Mammalian cell culture and plasmids. HeLa cells were cultured according to standard procedures. HeLa S3 cells were maintained in DMEM/F12 (Gibco), supplemented with 10% FCS and L-glutamine. Confluent cells were transfected using Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. Cells were treated with MG132 (or DMSO control) 24 h after transfection and analysed 4–8 h later by microscopy. HeLa cells were split onto coverslips, washed twice with PBS, fixed with 4% paraformaldehyde (in PBS) for 20 min and washed twice with PBS. GFP–Ubc9[ts], GFP–Ubc9[WT] and CHFP–VHL were sub-cloned from their original pESC vectors into pcDNA3.1 (Invitrogen) vectors.

Fluorescence microscopy. Conventional epifluorescence micrographs were obtained from live yeast cells on a Zeiss Axiovert microscope with a ×100 oil lens (NA1.4; Zeiss). Digital (12-bit) images were acquired with a cooled CCD (Princeton Instruments) and processed by using METAMORPH software (Universal Imaging). For deconvolution microscopy, yeast cells were fixed on glass coverslips in 4% paraformaldehyde. Deconvoluted images were acquired by using an Olympus microscope. Digital images (12 bit) were digitally deconvoluted by using DELTAVISION hardware and software (Applied Precision). Live-cell imaging was performed using the Marianas system from Intelligent Imaging Innovations equipped with the MicroPoint FRAP laser system (Photonic Instruments).

VHL solubility and ubiquitination assay. Yeast were grown, collected and lysed according to standard protocols[25]. Cells expressing VHL were grown at 30 °C or 37 °C, collected, washed once with sterile double-distilled water, and resuspended in 1× native yeast lysis buffer (30 mM HEPES (pH 8.0), 150 mM NaCl, 1 % glycerol, 1 mM DTT, 1 mM PMSF and 1 μg ml^{-1} pepstatin-A; and 1 mM NEM for ubiquitination assays). Where indicated, lysis buffer also contained 0.5% Triton. Pellets were frozen in liquid nitrogen and lysates were prepared by beating in liquid nitrogen (3 min) and clarified by centrifugation at 6,000*g* for 5 min at 4 °C. Fifty microlitres of this supernatant was set aside as total protein. Fifty microlitres was spun at 16,000*g* for 30 min at 4 °C. This supernatant was removed and designated the soluble fraction. The pellet was resolubilized by heating in 50 μl 1 × SDS sample buffer. Fifty microlitres of 4 × SDS sample buffer was added to the total protein and soluble fraction samples. Equal amounts of each fraction were resolved by SDS–PAGE followed by immunoblot analysis with anti-GFP or anti-Myc antisera. For gel aggregation assays equivalent total protein amounts of lysate were run on SDS–PAGE and both stacking and resolving gels were transferred and analysed by immunoblot as described in ref 17.

Fluorescence microscopy. Conventional epifluorescence micrographs were obtained from live yeast cells on a Zeiss Axiovert microscope with a ×100 oil lens (NA1.4). Digital (12-bit) images were acquired with a cooled CCD (Princeton Instruments) and processed by using METAMORPH software (Universal Imaging). The excitation filters used for conventional microscopy were 500DF20 (GFP), 540DF20 (Rhodamine) and 570DF20 (Texas red). Emission filters were 535DF20 (GFP), 560DF20 (Rhodamine) and 630DF25 (Texas red). The dichroics were: 505 DCLP (GFP) and 595 DCLP (Texas red).

For deconvolution microscopy, yeast cells were fixed on glass coverslips in 4 % paraformaldehyde. Deconvoluted images were acquired by using an Olympus microscope with 436 DF10 (CFP) and 500DF20 (YFP) filters for excitation and 470 DF30 (CFP) and 535 DF30 (YFP) filters for emission. Digital images (12-bit) were digitally deconvoluted by using DELTAVISION hardware and software (Applied Precision). Live-cell imaging was performed using the Marianas system from Intelligent Imaging Innovations equipped with the MicroPoint FRAP laser system (Photonic Instruments).

Electron microscopy. Cells were fixed and processed as described in ref. 55. Briefly, 25 ml cultures of exponentially growing cells (5×10^6 cells ml^{-1}) in minimal medium were quickly harvested by vacuum filtration over a 0.45 μm nitrocellulose membrane; filtration was stopped when the total volume in the filter apparatus was < 5 ml, but not dry. To this concentrated cell suspension, still on the filter membrane, 15 ml of freshly prepared, room temperature fixative 40 mM potassium phosphate, pH 6.7, 0.5 M sorbitol, 4% formaldehyde freshly prepared from paraformaldehyde (Polysciences), 0.2% glutaraldehyde (EM grade, Polysciences), 1 mM $MgCl_2$, and 1 mM EGTA, pH 8, was added and mixed rapidly with the cells by pipetting the suspension several times. The cell suspension was then transferred to a 50 ml polypropylene centrifuge tube and incubated at room temperature for approximately 1 h.

The fixed cells were then centrifuged at low speed in a clinical centrifuge and the pellet was resuspended in 40 mM potassium phosphate buffer (pH 6.7) containing 0.25 M sorbitol and transferred to Eppendorf tubes. The cells were again centrifuged and washed in 40 mM potassium phosphate buffer (pH 6.7). The final pellet of fixed cells was resuspended in 1 ml 1% sodium metaperiodate to make the cell wall more permeable, incubated for 10 min at room temperature, and then centrifuged and resuspended in 1 ml distilled water. Next, to block free aldehyde groups, the cells were centrifuged, resuspended in 1 ml 50 mM ammonium chloride and incubated for 10 min at room temperature.

The cells were then washed once in distilled water, centrifuged at low speed and immediately dehydrated (on ice) by resuspending the cell pellet in 70% (v/v) ice-cold ethanol and incubating on ice for 5 min. The cells were similarly centrifuged and sequentially resuspended in 80%, 85%, 90%, 95% ice-cold ethanol and finally once in 100% ice-cold ethanol. A final dehydration and centrifugation in 100% ethanol at room temperature was performed twice. The dehydrated cells then were infiltrated with room temperature L. R. White resin (Polysciences) and prepared for polymerization as described[56] except that infiltration of resin into the cells was done without application of vacuum and harvesting of cells was by centrifugation. The resin was polymerized by incubation at 47 °C for approximately 48 h.

Thin sections measuring approximately 60–70 nm (as determined by a grey/silver interference colour) were cut with a diamond knife and were picked up on 300 mesh nickel grids (Polysciences), which had been made sticky with a dilute formvar solution[56].

Affinity purified rabbit antibodies directed against GFP were a gift to J. Mulholland (CSIF) from P. Silver's laboratory (Harvard University). The secondary antibodies used were 10 nm gold-conjugated, anti-rabbit IgG (goat) secondary antibodies (BioCell). Antibody incubations were performed as described previously[55]. Primary and secondary antibodies were diluted 1:50 in PBST (140 mM NaCl, 3 mM KCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 0.05% Tween 20) containing 0.5 % BSA (bovin serum albumin) and 0.5 % ovalbumin (Sigma) and were incubated at room temperature for 1–2 h, with cell sections mounted on grids as described above. In the absence of the primary antibody, the anti-rabbit secondary antibodies did not react with the cell sections. After immunolocalization cell sections were post-fixed and stained with uranyl acetate and lead citrate as previously described [55]. All observations were made on a JEOL 1230 transmission electron microscope at an accelerating voltage of 80 kV using a 20-μm-diameter objective aperture using a Gatan 967 cooled CCD camera for image acquisition.

51. Prinz, A., Hartmann, E. & Kalies, K. U. Sec61p is the main ribosome receptor in the endoplasmic reticulum of *Saccharomyces cerevisiae*. *Biol. Chem.* **381,** 1025–1029 (2000).
52. Ghislain, M., Udvardy, A. & Mann, C. *S. cerevisiae* 26S protease mutants arrest cell division in G2/metaphase. *Nature* **366,** 358–362 (1993).
53. Chen, P., Johnson, P., Sommer, T., Jentsch, S. & Hochstrasser, M. Multiple ubiquitin-conjugating enzymes participate in the *in vivo* degradation of the yeast MAT α 2 repressor. *Cell* **74,** 357–369 (1993).
54. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285,** 901–906 (1999).
55. Mulholland, J. *et al.* Ultrastructure of the yeast actin cytoskeleton and its association with the plasma membrane. *J Cell Biol.* **125,** 381–391 (1994).
56. Wright, R. & Rine, J. Transmission electron microscopy and immunocytochemical studies of yeast: analysis of HMG-CoA reductase overproduction by electron microscopy. *Methods in cell biology* **31,** 473–512 (1989).

LETTERS

A common mass scale for satellite galaxies of the Milky Way

Louis E. Strigari[1], James S. Bullock[1], Manoj Kaplinghat[1], Joshua D. Simon[2], Marla Geha[3], Beth Willman[4] & Matthew G. Walker[5]

The Milky Way has at least twenty-three known satellite galaxies that shine with luminosities ranging from about a thousand to a billion times that of the Sun. Half of these galaxies were discovered[1,2] in the past few years in the Sloan Digital Sky Survey, and they are among the least luminous galaxies in the known Universe. A determination of the mass of these galaxies provides a test of galaxy formation at the smallest scales[3,4] and probes the nature of the dark matter that dominates the mass density of the Universe[5]. Here we use new measurements of the velocities of the stars in these galaxies[6,7] to show that they are consistent with them having a common mass of about $10^7 M_\odot$ within their central 300 parsecs. This result demonstrates that the faintest of the Milky Way satellites are the most dark-matter-dominated galaxies known, and could be a hint of a new scale in galaxy formation or a characteristic scale for the clustering of dark matter.

Many independent lines of evidence strongly argue for the presence of dark matter in galaxies, in clusters of galaxies, and throughout the observable Universe[5]. Its identity, however, remains a mystery. The gravity of dark matter overwhelms that of the normal atoms and molecules and hence governs the formation and evolution of galaxies and large-scale structure[8–10]. In the currently favoured models of dark matter, structure in the Universe forms hierarchically, with smaller gravitationally bound clumps of dark matter—haloes—merging to form progressively larger objects.

The mass of the smallest dark matter halo is determined by the particle properties of dark matter. Dark matter candidates characterized as cold dark matter can form haloes that are many orders of magnitude smaller than the least luminous haloes that we infer from observations. Cosmological simulations of cold dark matter predict that galaxies like the Milky Way should be teeming with thousands of dark matter haloes with masses $\sim 10^6 M_\odot$, with a steadily increasing number as we go to the smallest masses[11–14]. A large class of dark matter candidates characterized as 'warm' would predict fewer of these small haloes[15]. However, even for cold dark matter it is uncertain what fraction of the small dark matter haloes should host visible galaxies, as the ability of gas to cool and form stars in small dark matter haloes depends on a variety of poorly understood physical processes[16–20].

The smallest known galaxies hosted by their own dark matter haloes are the dwarf spheroidal satellites of the Milky Way[3,4]. These objects have very little gas and no signs of recent star formation. The least luminous galaxies were recently discovered in the Sloan Digital Sky Survey (SDSS)[1,2] and follow-up observations have revealed them to be strongly dominated by dark matter[6,21,22].

We have compiled line-of-sight velocity measurements of individual stars in 18 of the 23 known dwarf galaxies in the Milky Way[6,7]. We use these measurements to determine the dynamical mass of their dark matter haloes using a maximum likelihood analysis[23]. The dynamical mass is best constrained within the stellar extent, which corresponds to an average radius of $\sim$0.3 kiloparsecs (kpc) for all the satellites. We determine this mass, $M_{0.3}$, by marginalizing over a five-parameter density profile for dark matter that allows for both steep density cusps and flat cores in the central regions. It is important to note that the observed velocity dispersion of stars is determined by both the dynamical mass and the average anisotropy of the velocity dispersion (that is, difference between tangential and radial dispersion). The anisotropy is unknown and hence we marginalize over a three-parameter anisotropy function for stellar velocity that allows us to explore a range of orbital models for the stars[23].

Figure 1 shows the resulting determination of $M_{0.3}$. We find that all 18 dwarf galaxies are consistent with having a dynamical mass of 10^7 solar masses within 0.3 kpc of their centres, despite the fact that they have luminosity differences over four orders of magnitude. This result implies a central density for dark matter of $\sim 0.1 M_\odot\,\mathrm{pc}^{-3}$ in these galaxies. Earlier studies suggested that the highest luminosity dwarf galaxies all shared a common mass[4,24]. With larger stellar data sets, more than double the number of dwarf galaxies, and more detailed mass modelling, our results confirm this suggestion and conclusively establish that the dwarf galaxies of the Milky Way share a common mass scale.

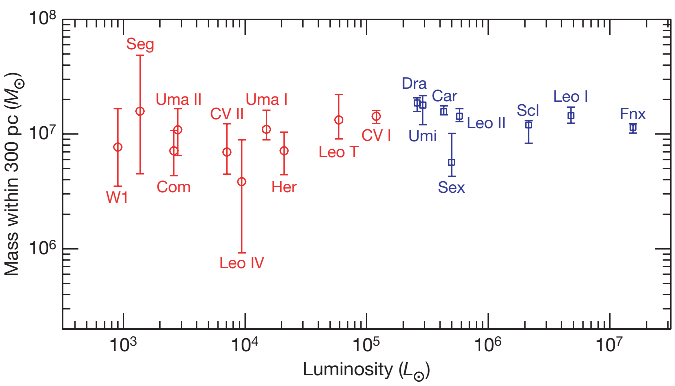


Figure 1 | The integrated mass of the Milky Way dwarf satellites, in units of solar masses, within their inner 0.3 kpc as a function of their total luminosity, in units of solar luminosities. The circle (red) points on the left refer to the newly discovered SDSS satellites, whereas the square (blue) points refer to the classical dwarf satellites discovered pre-SDSS. The error bars reflect the points where the likelihood function falls off to 60.6% of its peak value.

[1]Center for Cosmology, Department of Physics and Astronomy, University of California, Irvine, California 92697-4574, USA. [2]Department of Astronomy, California Institute of Technology, 1200 East California Boulevard, MS105-24, Pasadena, California 91125, USA. [3]Department of Astronomy, Yale University, PO Box 208101, New Haven, Connecticut 06520-8101, USA. [4]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [5]Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.

Because of the proximity of the dwarf galaxies to the Milky Way, it is possible that tidal effects could change the velocities of stars and thus affect the mass measurements. In the kinematic data, tidal forces could be revealed as a velocity gradient across the observed plane of the dwarf[25,26]. We have tested the dwarf galaxies for velocity gradients and have found no conclusive evidence of tidal effects (see Supplementary Information).

We fitted a $M_{0.3}$-luminosity relation to these data and obtained $M_{0.3} \propto L^{0.03 \pm 0.03}$. This result does not change significantly if we use the luminosity contained within 0.3 kpc rather than the total luminosity. The common mass scale of $\sim 10^7 M_\odot$ may thus reflect either a plummeting efficiency for galaxy formation at this mass scale, or the fact that dark matter haloes with lower masses simply do not exist.

The characteristic density of $0.1 M_\odot\ \mathrm{pc}^{-3}$ may be associated with a characteristic halo formation time. In theories of hierarchical structure formation, the central density of dark matter haloes is proportional to the mean density of matter in the Universe when the halo formed. The earlier the formation, the higher the density. For cold dark matter models, our measurement implies that these haloes collapsed at a redshift greater than about 12, or earlier than 100 million years after the Big Bang. Measurements of the cosmic microwave background[27] suggest that the Universe went from being neutral to ionized at redshift 11 ± 1.4. These dark matter haloes thus formed at roughly the same time that the Universe was re-ionized.

Within the context of the cold dark matter theory, high-resolution cosmological simulations can be used to relate the mass within the central regions of the dark matter halo to the depth of the gravitational potential well[28]. Simulations show that $M_{0.3} \approx 10^7 M_\odot \times (M_{\mathrm{total}}/10^9 M_\odot)^{0.35}$, where M_{total} is the mass of the halo before it was accreted into the Milky Way host potential. Thus, it is possible that the implied total mass scale of $10^9 M_\odot$ reflects the characteristic scale at which supernova feedback[29] or the imprint of the re-ionization of the Universe[18,30] could sharply suppress star formation.

Perhaps a more speculative, but certainly no less compelling, explanation of the common mass scale is that dark matter haloes do not exist with $M_{0.3}$ below $\sim 10^7 M_\odot$. This implies that these dwarf galaxies inhabit the smallest dark matter haloes in the Universe. Warm dark matter has a larger free-streaming length than standard cold dark matter, which implies that density perturbations are erased below a characteristic length scale, resulting in a higher minimum mass for dark matter haloes. A thermal warm dark matter candidate with mass of about 1 keV would imply a minimum halo mass of $10^9 M_\odot$. Thus, our mass determinations rule out thermal warm dark matter candidates with masses less than about 1 keV, but dark matter masses somewhat larger than 1 keV would yield a minimum dark matter halo mass consistent with the mass scale we observe.

Future imaging surveys of stars in the Milky Way will provide a more complete census of low-luminosity Milky Way satellites, with the prospects of determining whether astrophysics or fundamental dark matter physics is responsible for setting the common mass scale. In particular, the masses for the faintest dwarf galaxies will become more strongly constrained with more line-of-sight velocity data. This will sharpen the observational picture of galaxy formation on these small scales and provide data around which theories of galaxy formation may be built.

Note added in proof: While this paper was in press, the luminosities of several newly discovered dwarf satellites were updated[31].

Received 18 April; accepted 24 June 2008.

1. Willman, B. *et al.* A new Milky Way companion: unusual globular cluster or extreme dwarf satellite? *Astron. J.* **129,** 2692–2700 (2005).
2. Belokurov, V. *et al.* Cats and dogs, hair and a hero: a quintet of new Milky Way companions. *Astrophys. J.* **654,** 897–906 (2007).
3. Mateo, M. L. Dwarf galaxies of the Local Group. *Annu. Rev. Astron. Astrophys.* **36,** 435–506 (1998).
4. Gilmore, G. *et al.* The observed properties of dark matter on small spatial scales. *Astrophys. J.* **663,** 948–959 (2007).
5. Spergel, D. N. *et al.* Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: implications for cosmology. *Astrophys. J.* **170,** S377–S408 (2007).
6. Simon, J. D. & Geha, M. The kinematics of the ultra-faint Milky Way satellites: solving the missing satellite problem. *Astrophys. J.* **670,** 313–331 (2007).
7. Walker, M. G. *et al.* Velocity dispersion profiles of seven dwarf spheroidal galaxies. *Astrophys. J.* **667,** L53–L56 (2007).
8. Peebles, P. J. E. Large-scale background temperature and mass fluctuations due to scale invariant primeval perturbations. *Astrophys. J.* **263,** L1–L5 (1982).
9. White, S. D. M., Frenk, C. S. & Davis, M. Clustering in a neutrino-dominated universe. *Astrophys. J.* **274,** L1–L5 (1983).
10. Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. Formation of galaxies and large-scale structure with cold dark matter. *Nature* **311,** 517–525 (1984).
11. Klypin, A., Kravtsov, A. V., Valenzuela, O. & Prada, F. Where are the missing galactic satellites? *Astrophys. J.* **522,** 82–92 (1999).
12. Moore, B. *et al.* Dark matter substructure within galactic halos. *Astrophys. J.* **524,** L19–L22 (1999).
13. Diemand, J., Moore, B. & Stadel, J. Earth-mass dark matter haloes as the first structures in the early Universe. *Nature* **433,** 389–391 (2005).
14. Diemand, J., Kuhlen, M. & Madau, P. Dark matter substructure and gamma-ray annihilation in the Milky Way halo. *Astrophys. J.* **657,** 262–270 (2007).
15. Bode, P., Ostriker, J. P. & Turok, N. Halo formation in warm dark matter models. *Astrophys. J.* **556,** 93–107 (2001).
16. Efstathiou, G. Suppressing the formation of dwarf galaxies via photoionization. *Mon. Not. R. Astron. Soc.* **256,** 43P–47P (1992).
17. Kauffmann, G., White, S. D. M. & Guiderdoni, B. The formation and evolution of galaxies within merging dark matter haloes. *Mon. Not. R. Astron. Soc.* **264,** 201–218 (1993).
18. Bullock, J. S., Kravtsov, A. V. & Weinberg, D. H. Reionization and the abundance of galactic satellites. *Astrophys. J.* **539,** 517–521 (2000).
19. Kravtsov, A. V., Gnedin, O. Y. & Klypin, A. A. The tumultuous lives of galactic dwarfs and the missing satellites problem. *Astrophys. J.* **609,** 482–497 (2004).
20. Mayer, L., Kazantzidis, S., Mastropietro, C. & Wadsley, J. Early gas stripping as the origin of the darkest galaxies in the Universe. *Nature* **445,** 738–740 (2007).
21. Muñoz, R. R. *et al.* Exploring halo substructure with giant stars: the dynamics and metallicity of the dwarf spheroidal in Boötes. *Astrophys. J.* **650,** L51–L54 (2006).
22. Martin, N. F., Ibata, R. A., Chapman, S. C., Irwin, M. & Lewis, G. F. A. Keck/DEIMOS spectroscopic survey of faint Galactic satellites: searching for the least massive dwarf galaxies. *Mon. Not. R. Astron. Soc.* **380,** 281–300 (2007).
23. Strigari, L. E. *et al.* Redefining the missing satellites problem. *Astrophys. J.* **669,** 676–683 (2007).
24. Mateo, M., Olszewski, E. W., Pryor, C., Welch, D. L. & Fischer, P. The Carina dwarf spheroidal galaxy: How dark is it? *Astron. J.* **105,** 510–526 (1993).
25. Piatek, S. & Pryor, C. The effect of galactic tides on the apparent mass-to-light ratios in dwarf spheroidal galaxies. *Astron. J.* **109,** 1071–1085 (1995).
26. Fellhauer, M. *et al.* Is Ursa Major II the progenitor of the orphan stream? *Mon. Not. R. Astron. Soc.* **375,** 1171–1179 (2007).
27. Dunkley, J. *et al.* Five-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: likelihoods and parameters from the WMAP data. Preprint available at ⟨http://arxiv.org/abs/0803.0586⟩ (2008).
28. Bullock, J. S. *et al.* Profiles of dark haloes: evolution, scatter and environment. *Mon. Not. R. Astron. Soc.* **321,** 559–575 (2001).
29. Dekel, A. & Silk, J. The origin of dwarf galaxies, cold dark matter, and biased galaxy formation. *Astrophys. J.* **303,** 39–55 (1986).
30. Wyithe, J. S. B. & Loeb, A. Suppression of dwarf galaxy formation by cosmic reionization. *Nature* **441,** 322–324 (2006).
31. Martin, N. F., de Jong, J. T. A. & Rix, H.-W. A comprehensive maximum likelihood analysis of the structural properties of faint Milky Way satellites. Preprint at ⟨http://arxiv.org/abs/0805.2945⟩ (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Johnston and S. White for discussion on this paper, and J. Strader for help in the acquisition of data for the Willman 1 satellite.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.E.S. (lstrigar@uci.edu).

LETTERS

Experimental demonstration of a BDCZ quantum repeater node

Zhen-Sheng Yuan[1,2]*, Yu-Ao Chen[1,2]*, Bo Zhao[1], Shuai Chen[1], Jörg Schmiedmayer[3] & Jian-Wei Pan[1,2]

Quantum communication is a method that offers efficient and secure ways for the exchange of information in a network. Large-scale quantum communication[1–4] (of the order of 100 km) has been achieved; however, serious problems occur beyond this distance scale, mainly due to inevitable photon loss in the transmission channel. Quantum communication eventually fails[5] when the probability of a dark count in the photon detectors becomes comparable to the probability that a photon is correctly detected. To overcome this problem, Briegel, Dür, Cirac and Zoller (BDCZ) introduced the concept of quantum repeaters[6], combining entanglement swapping[7] and quantum memory to efficiently extend the achievable distances. Although entanglement swapping has been experimentally demonstrated[8], the implementation of BDCZ quantum repeaters has proved challenging owing to the difficulty of integrating a quantum memory. Here we realize entanglement swapping with storage and retrieval of light, a building block of the BDCZ quantum repeater. We follow a scheme[9,10] that incorporates the strategy of BDCZ with atomic quantum memories[11]. Two atomic ensembles, each originally entangled with a single emitted photon, are projected into an entangled state by performing a joint Bell state measurement on the two single photons after they have passed through a 300-m fibre-based communication channel. The entanglement is stored in the atomic ensembles and later verified by converting the atomic excitations into photons. Our method is intrinsically phase insensitive and establishes the essential element needed to realize quantum repeaters with stationary atomic qubits as quantum memories and flying photonic qubits as quantum messengers.

Although the BDCZ protocol[6] attracted much interest as a solution to extending the communication length, the absence of quantum memory has hindered the implementation of quantum repeaters. In 2001, Duan, Lukin, Cirac and Zoller (DLCZ) proposed an alternative quantum repeater scheme[11] where linear optics and atomic ensembles are used to incorporate entanglement connection and quantum memory into a single unit. Motivated by the DLCZ protocol, number-state entanglement between two atomic ensembles has been observed[12,13]. Most recently, a functional quantum node[14] based on asynchronous preparation of number-state entanglement for two pairs of atomic ensembles—the basic element of the DLCZ protocol—has also been demonstrated.

However, two serious drawbacks make the DLCZ-type functional quantum nodes[11,14] unlikely to be a realistic solution for long-distance quantum communication[9,10,15]. First, the required long-term subwavelength stability of the path difference between two arms of a large-scale single-photon interferometer spanning the whole communication distance is very difficult to achieve[9,10], even with the latest and most sophisticated technology for coherent optical phase transfer[16]. Second, the swapping of number-state entanglement using a single-photon interferometer leads to the growth of a vacuum component in the generated state, and to the rapid growth of errors due to multiple emissions from individual ensembles[15].

A novel solution[9,10] is to combine the atomic quantum memory in DLCZ and the strategy of BDCZ. In this scheme, two-photon interference is used to generate long-distance entanglement, so the stability requirement for the path differences is determined by the coherence length of the photons and is consequently seven orders of magnitude looser[17] than in the DLCZ protocol. Moreover, the vacuum component can be suppressed and is no longer a dominant term after a few entanglement connections[9,10]. Following this scheme, we demonstrate here the implementation of a quantum repeater node, involving entanglement swapping with the function of storage and retrieval of light. A high precision of local operations has been achieved that surpasses the theoretical threshold[6] required for the realization of robust quantum repeaters for long-distance quantum communication.

In our experiment, to demonstrate entanglement swapping with storage and retrieval of light, we follow three steps: implementing two atom–photon entanglement sources, sending the flying qubits (the photons) to an intermediate station for a Bell state measurement (BSM), and verifying the entanglement between the stationary qubits (the two remote atomic ensembles).

Unlike previous atom–photon entanglement sources realized with trapped ions[18], single atoms in a cavity[19], or two spatially separated atomic ensembles[20], we use here two collective excitations in different spatial modes of a single atomic ensemble to implement the atom–photon entanglement[21]. In contrast to the method in which two separated spatial regions in one atomic cloud are covered by their own 'write' and 'read' beams[14], here the two excitation modes share the same write and read beams, which offers high-quality entanglement and long-term stability.

The basic principle is shown in Fig. 1 (see Methods). Alice and Bob each have a cold atomic ensemble consisting of about 10^8 atoms of ^{87}Rb loaded by magneto-optical traps (MOTs). At each site, atoms are first prepared in the initial state $|a\rangle$, followed by a weak write pulse. Two anti-Stokes fields AS_L and AS_R induced by the write beam via spontaneous Raman scattering are collected at $\pm 3^\circ$ relative to the propagating direction of the write beam. This defines two spatial modes of excitation in the atomic ensembles (L and R), which constitute our memory qubit.

The two anti-Stokes fields in modes L and R are adjusted to have equal excitation probability and orthogonal polarizations. The two fields are then overlapped at a polarizing beam splitter PBS2 (Fig. 1) and coupled into a single-mode fibre. Neglecting the vacuum state and higher order excitations, the entangled state between the atomic

[1]Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Philosophenweg 12, 69120 Heidelberg, Germany. [2]Hefei National Laboratory for Physical Sciences at Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei, Anhui 230026, China. [3]Atominstitut der Österreichischen Universitäten, TU-Wien, A-1020 Vienna, Austria.
*These authors contributed equally to this work.

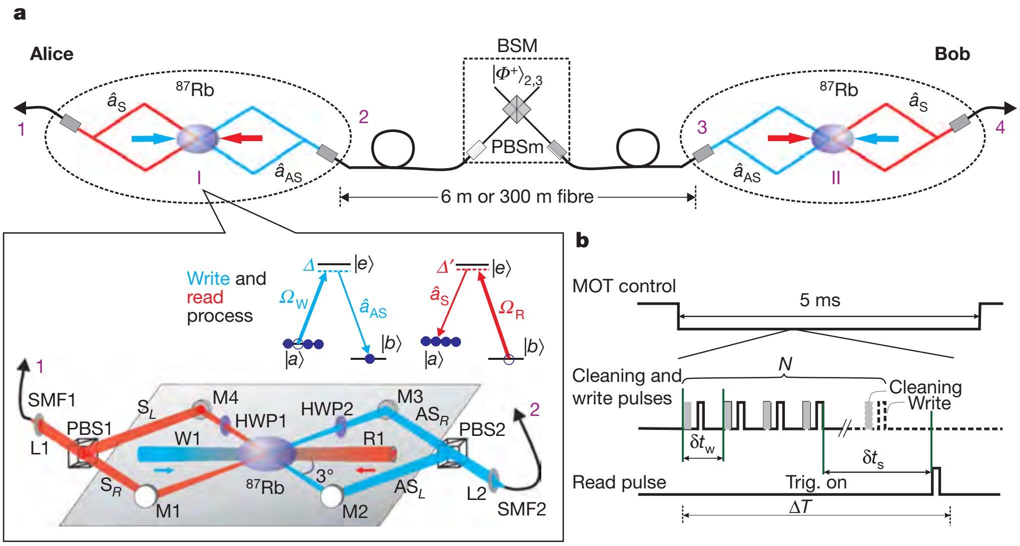


Figure 1 | The experimental scheme for entanglement swapping. **a**, Photons 2 and 3 overlap at BSM (Bell state measurement) and are projected to the state of $|\Phi^+\rangle_{2,3}$ through which the entanglement is generated between the two atomic ensembles I and II confined by magneto-optical traps (MOTs) in different glass cells separated by ~60 cm. Here 1, 2, 3 and 4 indicate four photons emerging from the anti-Stokes modes ($\hat{a}_{AS}$) and Stokes modes ($\hat{a}_S$), and $|\Phi^+\rangle_{2,3}$ is one of the four Bell states. Inset, atom–photon entanglement. Shown are energy levels $\{|a\rangle,|b\rangle,|e\rangle\} = \{|5S_{1/2},F=2\rangle, |5S_{1/2},F=1\rangle, |5P_{1/2},F=2\rangle\}$ and the configuration of light beams. PBS, polarizing beam splitter; HWP, half-wave plate; L, lens; M, mirror; SMF, single-mode fibre; W, write beam; R, read beam; S, Stokes field; AS, anti-Stokes field; Ω, Rabi frequency of light fields. **b**, The time sequence of the experimental procedure at each site. For 6-m (300-m) fibre connection, there are 250 (200) experiment cycles in 5 ms and ΔT is 16 μs (20 μs) for one cycle, which contains $N = 10$ ($N = 8$) write sequences. The interval between two neighbouring write pulses is $\delta t_w = 1$ μs (1.5 μs) and δt_s is the storage time. Whenever there is a desired coincidence event between photons 2 and 3, the following write sequence is stopped by a feedback circuit and the retrieve process can be started (at the time point labelled 'Trig. on').

and photonic qubits can be described effectively as

$$|\Psi\rangle_{\mathrm{at-ph}} = \frac{1}{\sqrt{2}}\left(|H\rangle|R\rangle + e^{i\phi_1}|V\rangle|L\rangle\right) \quad (1)$$

where $|H\rangle/|V\rangle$ denotes horizontal/vertical polarization of the single anti-Stokes photon, $|L\rangle/|R\rangle$ denotes a single collective excitation in ensemble L/R, and ϕ_1 is the propagating phase difference between the two anti-Stokes fields before they overlap at PBS2. The atom–photon entangled state (equation (1)) is equivalent to the maximally polarization-entangled state generated by spontaneous parametric down-conversion[22].

In this way, one can implement two separate and remote atom–photon entanglement sources at Alice's site (I) and Bob's site (II), respectively. To make the higher order excitations negligible, a low excitation probability ($\chi_m \approx 0.01$) is chosen for the collective modes m ($m = L, R$). Owing to the imperfect coupling of light modes, the transmission loss, and the inefficiency of single-photon detectors, the overall detection efficiency of an emerging anti-Stokes photon (η_{AS}) is around 25%. To check the quality of atom–photon entanglement, a read pulse (see Methods) is applied after a controllable time delay δt_s to convert the atomic collective excitation back into a Stokes field. Ideally, the retrieve efficiency of the Stokes fields should reach unity. However, various imperfections, such as low optical depth of the atomic ensembles and mode mismatching between the write and read pulses, lead to a 35% retrieve efficiency. Together with the non-ideal collection and detection efficiency (~40%) of single-photon detectors, the overall detection efficiency of the Stokes photon is around 15%. After combining the two retrieved Stokes fields on PBS1 (see Fig. 1), the anti-Stokes and Stokes fields are in the following maximally polarization-entangled state

$$|\Psi\rangle_{\mathrm{AS,S}} = \frac{1}{\sqrt{2}}\left(|H\rangle_{\mathrm{AS}}|H\rangle_{\mathrm{S}} + e^{i(\phi_1+\phi_2)}|V\rangle_{\mathrm{AS}}|V\rangle_{\mathrm{S}}\right) \quad (2)$$

where ϕ_2 represents the propagating phase difference between two Stokes fields before they overlap at PBS1. In our experiment, the total phase $\phi_1 + \phi_2$ is actively stabilized via a built-in Mach–Zehnder interferometer and fixed to zero (see Supplementary Information). With a time delay $\delta t_s = 0.5$ μs, the measured polarization correlations of the Stokes and anti-Stokes photons show a strong violation of a CHSH-type Bell's inequality, with a visibility of 92%. Further measurement shows our atom–photon entanglement still survives up to a storage time of $\delta t_s = 20$ μs (see Supplementary Information).

We now describe the entanglement generation between atomic ensembles I and II via entanglement swapping. As shown in Fig. 1, photon 2 from Alice and photon 3 from Bob are both sent through a 3-m optical fibre to an intermediate station for a joint BSM. In the experiment, we chose to analyse the projection onto the Bell state $|\Phi^+\rangle_{2,3} = (1/\sqrt{2})(|H\rangle_2|H\rangle_3 + |V\rangle_2|V\rangle_3)$, which is achieved by overlapping photons 2 and 3 onto a polarizing beam splitter (PBSm) and performing a proper polarization decomposition in the output modes and a subsequent coincidence detection[23]. Conditioned on detecting a $|\Phi^+\rangle_{2,3}$ state at the intermediate station, the two remote atomic ensembles are projected onto an identical entangled state[7,8]: $|\phi^+\rangle_{\mathrm{I,II}} = (1/\sqrt{2})(|L\rangle_{\mathrm{I}}|L\rangle_{\mathrm{II}} + |R\rangle_{\mathrm{I}}|R\rangle_{\mathrm{II}})$.

It is noteworthy that double excitations in either atomic ensemble I or II will cause false events in the BSM[9,10], which reduce the success probability of entanglement swapping by a factor of 2. Experimentally, the false events can be eliminated at the stage of entanglement verification by the fourfold coincidence measurement of photons 1, 2, 3 and 4. Note that the detection time of photons 1 and 4 is later than that of photons 2 and 3 by an interval δt_s, the storage time in quantum memories. More importantly, such false events do not affect the applications of our experimental method in quantum repeaters, since the generation of entanglement will be deterministic after a second step of connecting two such nodes, where double excitations are excluded automatically[9,10].

The entanglement established between atomic ensembles I and II can be verified by converting the atomic spins into an entangled photon pair 1 and 4, which are in the state $|\Phi^+\rangle_{1,4}$. Here we measure the S parameter in a CHSH-type Bell's inequality,

$$S = \left|E(\theta_1,\theta_4) - E(\theta_1,\theta_4') - E(\theta_1',\theta_4) - E(\theta_1',\theta_4')\right| \quad (3)$$

where $E(\theta_1, \theta_4)$ is the correlation function, and θ_1 and θ_1' (θ_4 and θ_4')

are the measured polarization bases of photon 1 (4). In the measurement, the polarization settings are (0°, 22.5°), (0°, −22.5°), (45°, 22.5°) and (45°, −22.5°), respectively.

At a storage time $\delta t_s = 500$ ns, the measured correlation functions (shown in Fig. 2) result in $S = 2.26 \pm 0.07$, which violates Bell's inequality by 3 standard deviations. To observe the lifetime of the entanglement between two remote memory qubits, we measure the interference visibility of photons 1 and 4 as a function of the storage time by choosing the polarization basis of $+/-$ (shown in Fig. 3, with $|+\rangle = (1/\sqrt{2})(|H\rangle + |V\rangle)$ and $|-\rangle = (1/\sqrt{2})(|H\rangle - |V\rangle)$). Up to a storage time of 4.5 μs, the visibility is still higher than the threshold of $1/\sqrt{2}$, sufficient for the violation of Bell's inequality. From the visibilities of the atom–photon and atom–atom entanglements, the precision of local operations at the BSM station is estimated to be better than 97% (see Supplementary Information). We emphasize that this precision achieved here surpasses the threshold of 95% for local operations of independent photons necessary for future entanglement purification and connections[6], and therefore fits the requirement for a scalable quantum network.

To demonstrate the robustness of our protocol in the generation of quantum entanglement between two atomic ensembles over large distances, we changed the length of the two connecting fibres from 3 m to 150 m. The anti-Stokes photon is delayed 730 ns and the connection length between Alice and Bob is 300 m. The entanglement swapping can be quantified by the fidelity of the measured state of the atomic ensembles. To determine the fidelity, we write the density matrix of $|\phi^+\rangle_{\mathrm{I,II}}$ in terms of the Pauli matrices:

$$|\phi^+\rangle\langle\phi^+|_{\mathrm{I,II}} = \frac{1}{4}\left(I + \hat{\sigma}_x\hat{\sigma}_x - \hat{\sigma}_y\hat{\sigma}_y + \hat{\sigma}_z\hat{\sigma}_z\right) \qquad (4)$$

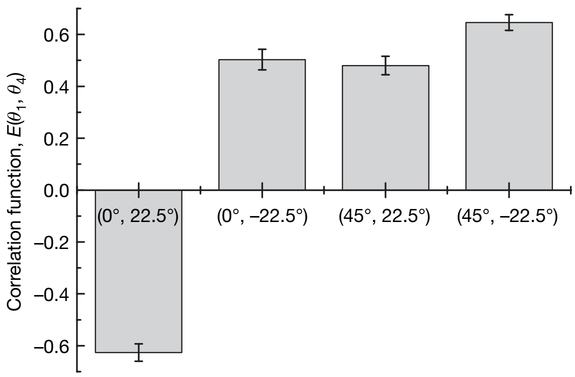


Figure 2 | Correlation functions of a CHSH-type Bell's inequality with a storage time δt_s = 500 ns. The x axis of the bar graph is labelled with (θ_1, θ_4), where θ_1 and θ_4 are the measured polarization bases of photons 1 and 4, respectively. Error bars represent statistical errors, which are ±1 s.d.

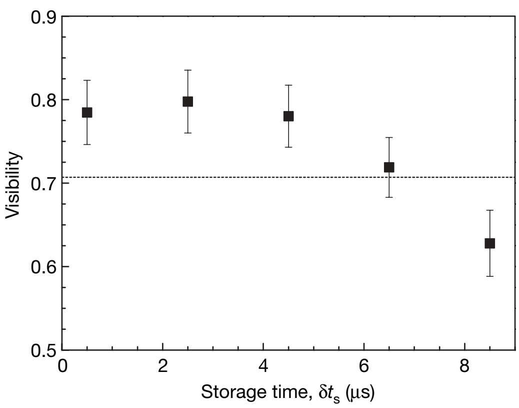


Figure 3 | Visibility of the atom–atom entanglement as a function of the storage time with 6-m fibre connection. Filled black squares, visibility; dotted line, threshold for the violation of the CHSH-type Bell's inequality. Error bars represent statistical errors, which are ±1 s.d.

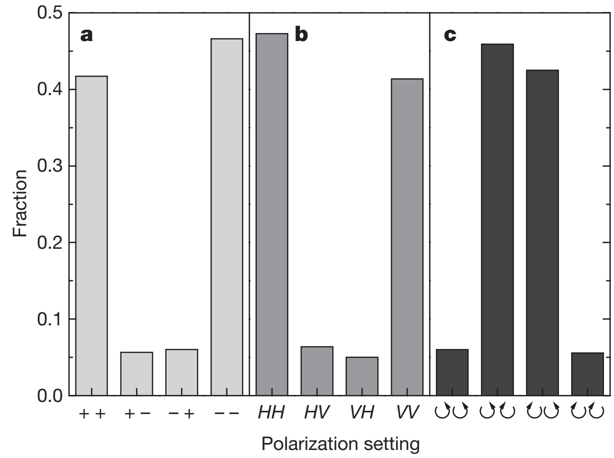


Figure 4 | Polarization analysis of photons 1 and 4 when the connection channel is a 300-m fibre. The bars indicate the observed fractions for the counts of specific polarization observables over the total counts of all the polarization observables. For instance, the bar at ++ shows the ratio between the counts of the observable $|++\rangle\langle++|$ and the total counts of observables $|++\rangle\langle++|$, $|+-\rangle\langle+-|$, $|-+\rangle\langle-+|$ and $|--\rangle\langle--|$. The polarization bases are chosen as **a**, $+/-$; **b**, H/V; **c**, $\circlearrowleft/\circlearrowright$.

Here $\sigma_x = |+\rangle\langle+| - |-\rangle\langle-|$, $\sigma_y = |\circlearrowright\rangle\langle\circlearrowright| - |\circlearrowleft\rangle\langle\circlearrowleft|$ and $\sigma_z = |H\rangle\langle H| - |V\rangle\langle V|$, with $|\circlearrowleft\rangle = (1/\sqrt{2})(|H\rangle + i|V\rangle)$ and $|\circlearrowright\rangle = (1/\sqrt{2})(|H\rangle - i|V\rangle)$. After a storage time of 1,230 ns (with a 730-ns delay being taken into account), the two retrieved photons 1 and 4 are sent to their own polarization analyser. Three series of polarization settings are used and the measured local observables are shown in Fig. 4. The fidelity of final state ρ_{exp} on $|\phi^+\rangle$ is given by $F = \mathrm{Tr}\left(\rho_{\mathrm{exp}}|\phi^+\rangle_{\mathrm{I,II}}\langle\phi^+|\right) = 0.83 \pm 0.02$, with 2.5 standard deviations beyond the threshold of 0.78 to violate the CHSH-type Bell's inequality for Werner states, demonstrating the success of entanglement swapping. This fidelity is comparable to the average value achieved in the DLCZ-type functional quantum node[14].

In summary, we have successfully demonstrated high-precision entanglement swapping with storage and retrieval of light, a building block for quantum repeaters. The extension of our work to longer chains will involve many quantum repeater nodes. To achieve this ambitious goal, several quantities—such as the lifetime and retrieve efficiency of the quantum memory, and the fidelity and generation rate of the entanglement state—still need to be improved significantly. We suggest three ways forward. First, better compensation of the residual magnetic field and trapping the atoms in 'clock states'[24] with a blue-detuned optical trap[25] should improve the lifetime to ~1 s. Second, a high optical density of the atomic cloud, achieved with the help of traps or by coupling the atoms into an optical cavity[26], should increase the retrieve efficiency close to unity. These improvements of the quantum memory would greatly enhance the fidelity and generation rate of the entanglement. Last, by local generation of entangled pairs of atomic excitations together with the present technique of entanglement swapping, the entanglement distribution rate can be greatly improved[27]. Not only does our work enable immediate experimental investigations of various quantum information protocols, but—with the above-mentioned future improvements—entanglement swapping with storage and retrieval of light would also open the way to long-distance quantum communication.

METHODS SUMMARY

As shown in Fig. 1, Alice and Bob each have a cold atomic ensemble consisting of about 10^8 atoms of ^{87}Rb with temperature ~100 μK. After 20 ms of loading atoms into their MOTs separated by ~60 cm, we switch off the laser beams and magnetic fields of the MOTs and start a 5-ms-long experiment cycle. At each site, atoms are first prepared in the initial state $|a\rangle$, followed by a (50 ns long,

$\sim$1 μW) weak write pulse, which has a beam waist of 240 μm and is 10 MHz red-detuned from the $|a\rangle \rightarrow |e\rangle$ transition. Two anti-Stokes fields AS_L and AS_R induced by the write beam via spontaneous Raman scattering are collected at $\pm 3^\circ$ relative to the propagating direction of the write beam (70 μm waist, $|e\rangle \rightarrow |b\rangle$). The excitation probability (χ_m) of the collective modes m ($m = L$, R) is low ($\chi_m \ll 1$); thus the state of the atom–photon field can be expressed as[11]

$$|\Psi\rangle_m \approx |0_{AS} 0_b\rangle_m + \sqrt{\chi_m} |1_{AS} 1_b\rangle_m + O(\chi_m)$$

and $|i_{AS} i_b\rangle_m$ denotes the i-fold excitation of the anti-Stokes field and the collective spin in the atomic ensemble. The read beam is counter-propagating and mode-matched with the write beam with a pulse length of 50 ns, a power of 60 μW and a frequency close to resonance of the $|b\rangle \rightarrow |e\rangle$ transition.

Received 9 March; accepted 3 July 2008.

1. Peng, C.-Z. *et al.* Experimental long-distance decoy-state quantum key distribution based on polarization encoding. *Phys. Rev. Lett.* **98,** 010505 (2007).
2. Rosenberg, D. *et al.* Long-distance decoy-state quantum key distribution in optical fiber. *Phys. Rev. Lett.* **98,** 010503 (2007).
3. Schmitt-Manderbach, T. *et al.* Experimental demonstration of free-space decoy-state quantum key distribution over 144 km. *Phys. Rev. Lett.* **98,** 010504 (2007).
4. Ursin, R. *et al.* Entanglement-based quantum communication over 144 km. *Nature Phys.* **3,** 481–486 (2007).
5. Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74,** 145–195 (2002).
6. Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: The role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81,** 5932–5935 (1998).
7. Żukowski, M., Zeilinger, A., Horne, M. A. & Ekert, A. K. "Event-ready-detectors" Bell experiment via entanglement swapping. *Phys. Rev. Lett.* **71,** 4287–4290 (1993).
8. Pan, J.-W., Bouwmeester, D., Weinfurter, H. & Zeilinger, A. Experimental entanglement swapping: Entangling photons that never interacted. *Phys. Rev. Lett.* **80,** 3891–3894 (1998).
9. Zhao, B., Chen, Z.-B., Chen, Y.-A., Schmiedmayer, J. & Pan, J.-W. Robust creation of entanglement between remote memory qubits. *Phys. Rev. Lett.* **98,** 240502 (2007).
10. Chen, Z.-B., Zhao, B., Chen, Y.-A., Schmiedmayer, J. & Pan, J.-W. Fault-tolerant quantum repeater with atomic ensembles and linear optics. *Phys. Rev. A* **76,** 022329 (2007).
11. Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414,** 413–418 (2001).
12. Matsukevich, D. N. & Kuzmich, A. Quantum state transfer between matter and light. *Science* **306,** 663–666 (2004).
13. Chou, C.-W. *et al.* Measurement-induced entanglement for excitation stored in remote atomic ensembles. *Nature* **438,** 828–832 (2005).
14. Chou, C.-W. *et al.* Functional quantum nodes for entanglement distribution over scalable quantum networks. *Science* **316,** 1316–1318 (2007).
15. Jiang, L., Taylor, J. M. & Lukin, M. D. Fast and robust approach to long-distance quantum communication with atomic ensembles. *Phys. Rev. A* **76,** 012301 (2007).
16. Foreman, S. M. *et al.* Coherent optical phase transfer over a 32-km fiber with 1 s instability at 10^{-17}. *Phys. Rev. Lett.* **99,** 153601 (2007).
17. Yuan, Z.-S. *et al.* Synchronized independent narrow-band single photons and efficient generation of photonic entanglement. *Phys. Rev. Lett.* **98,** 180503 (2007).
18. Moehring, D. L. *et al.* Entanglement of single-atom quantum bits at a distance. *Nature* **449,** 68–72 (2007).
19. Wilk, T., Webster, S. C., Kuhn, A. & Rempe, G. Single-atom single-photon quantum interface. *Science* **317,** 488–490 (2007).
20. Chen, Y.-A. *et al.* Memory-built-in quantum teleportation with photonic and atomic qubits. *Nature Phys.* **4,** 103–107 (2008).
21. Chen, S. *et al.* A stable atom-photon entanglement source for quantum repeaters. *Phys. Rev. Lett.* **99,** 180505 (2007).
22. Kwiat, P. G. *et al.* New high-intensity source of polarization-entangled photon pairs. *Phys. Rev. Lett.* **75,** 4337–4341 (1995).
23. Pan, J.-W. & Zeilinger, A. Greenberger-Horne-Zeilinger-state analyzer. *Phys. Rev. A* **57,** 2208–2211 (1998).
24. Harber, D. M., Lewandowski, H. J., McGuirk, J. M. & Cornell, E. A. Effect of cold collisions on spin coherence and resonance shifts in a magnetically trapped ultracold gas. *Phys. Rev. A* **66,** 053616 (2002).
25. Kuga, T. *et al.* Novel optical trap of atoms with a doughnut beam. *Phys. Rev. Lett.* **78,** 4713–4716 (1997).
26. Simon, J., Tanji, H., Thompson, J. K. & Vuletić, V. Interfacing collective atomic excitations and single photons. *Phys. Rev. Lett.* **98,** 183601 (2007).
27. Sangouard, N. *et al.* Robust and efficient quantum repeaters with atomic ensembles and linear optics. *Phys. Rev. A* **77,** 062301 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Dür for discussions. This work was supported by the Deutsche Forschungsgemeinschaft, the Alexander von Humboldt Foundation, and the European Commission through the Marie Curie Excellence Grant and the ERC Grant. This work was also supported by the National Fundamental Research Program (grant 2006CB921900), the CAS and the NNSFC.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Y.-A.C. (yuao@physi.uni-heidelberg.de) or J.-W.P. (jian-wei.pan@physi.uni-heidelberg.de).

LETTERS

Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO_2 levels

Daniel J. Lunt[1,2], Gavin L. Foster[3], Alan M. Haywood[4] & Emma J. Stone[1]

It is thought[1,2] that the Northern Hemisphere experienced only ephemeral glaciations from the Late Eocene to the Early Pliocene epochs (about 38 to 4 million years ago), and that the onset of extensive glaciations did not occur until about 3 million years ago[3,4]. Several hypotheses have been proposed to explain this increase in Northern Hemisphere glaciation during the Late Pliocene[5–11]. Here we use a fully coupled atmosphere–ocean general circulation model and an ice-sheet model to assess the impact of the proposed driving mechanisms for glaciation and the influence of orbital variations on the development of the Greenland ice sheet in particular. We find that Greenland glaciation is mainly controlled by a decrease in atmospheric carbon dioxide during the Late Pliocene. By contrast, our model results suggest that climatic shifts associated with the tectonically driven closure of the Panama seaway[5,6], with the termination of a permanent El Niño state[7–9] or with tectonic uplift[10] are not large enough to contribute significantly to the growth of the Greenland ice sheet; moreover, we find that none of these processes acted as a priming mechanism for glacial inception triggered by variations in the Earth's orbit.

During the Early and Middle Eocene (~56–40 Myr ago), when proxy data indicate high-latitude continental winter temperatures as high as 2 °C (ref. 12), Greenland was mostly free from ice[13]. There is some evidence of North Atlantic ice-rafted debris (IRD) in the Late Eocene and Early Oligocene (~38–30 Myr ago), indicating early ephemeral glaciations[1], and continued evidence for IRD through the Miocene[2]; however, the entry into an 'icehouse' world in the Northern Hemisphere, with extensive glaciation over much of Greenland, is thought to have occurred in the Late Pliocene around 3 Myr ago[3,4]. Several explanations have been proposed for this increase in glaciation during the Late Pliocene. Here we focus on their implications for the development of the Greenland ice sheet.

The 'Panama hypothesis'[5,6] suggests that the tectonically driven closure of the Panama seaway, between about 13 Myr and 2.5 Myr ago, led to an increase of the salinity contrast between the Pacific and Atlantic oceans, increased oceanic northward heat transport, and warmer and more evaporative surface water masses in the North Atlantic. This led to increased atmospheric moisture in the Arctic, greater snowfall over Greenland and ultimately an increase in ice volume. This hypothesis is supported by the coincidental timing of changes in salinity gradient between the Caribbean basin and the East Pacific, and pulses of increased IRD on the Icelandic Plateau[4].

The 'ENSO hypothesis' states that a 'permanent El Niño' state existed in the Early Pliocene and terminated in the Late Pliocene, evidence for which comes in the form of an increasing zonal gradient in sea surface temperature during this time period, reconstructed from the ratio Mg/Ca (ref. 7), oxygen isotope data ($\delta^{18}O$)[7] and alkenone data from sediment cores[8] in the eastern and western Equatorial Pacific. It has been suggested[9] that the transmission of El Niño-like anomalies from the tropics through large-scale planetary waves may have warmed the high latitudes of the Northern Hemisphere, and thus a permanent El Niño state may have acted to retard the onset of Northern Hemisphere glaciation. Therefore the loss of a permanent El Niño state in the Late Pliocene acted as a positive forcing mechanism for the onset of glaciation on Greenland.

The 'uplift hypothesis'[10] suggests that the Cenozoic uplift of the Rocky Mountains and Himalayas influenced atmospheric circulation, by causing larger Rossby wave amplitude and jet-stream deflection, to bring cooler air masses and increased moisture and snowfall over the incipient Greenland ice sheet.

The 'CO_2 hypothesis' states that decreased radiative forcing associated with a lowered concentration of atmospheric CO_2 led to cooler melt-season temperatures and decreased ablation, resulting in the net annual accumulation of snow necessary to grow an ice sheet. In contrast to the Quaternary, for which ice cores provide detailed and robust records[14] of past CO_2, pre-Quaternary atmospheric CO_2 concentration is a challenge to reconstruct owing to the lack of direct indicators. Estimates from the $^{13}C/^{12}C$ ratio of alkenones in marine sediments and boron isotopes in marine carbonates[15,16] suggest Eocene atmospheric CO_2 of the order of 1,000 p.p.m.v., falling to levels as low as 200 p.p.m.v. in the Middle Miocene, and then rising through the Miocene to levels similar to pre-industrial in the Early Pliocene[17]. Estimates from the $^{13}C/^{12}C$ ratio of marine organic matter[18] indicate mid-Pliocene (3.3–3.0 Myr ago) atmospheric CO_2 values of around 400 p.p.m.v., which is in broad agreement with estimates from the stomatal density of fossilized leaves[19] from the same period. The CO_2 hypothesis suggests that it is the fall from these mid-Pliocene values to lower values typical of the Quaternary that favoured the development of Northern Hemisphere glaciation and the growth of the Greenland ice sheet. It does not suggest what may have caused the variations in CO_2 through the Cenozoic; these variations may themselves have been ultimately tectonically driven[11], but this is distinct from the direct tectonic and orographic forcings of the Panama and uplift hypotheses.

It has also been suggested that none of the above hypotheses were capable of causing glacial inception on their own, but instead provided a 'priming' mechanism, with full glaciation triggered by variations in the Earth's orbit. This 'orbital-trigger hypothesis' is supported by the coincident timing of increased IRD with particularly large-amplitude oscillations on precessional timescales of boreal summer insolation[3].

To test the four main hypotheses—Panama, ENSO, uplift and CO_2—and the influence of orbital variations, we carried out a suite of simulations using a fully coupled atmosphere–ocean general circulation model (GCM), and used the modelled temperature and precipitation over Greenland to drive an off-line ice-sheet model. See Methods and Supplementary Information for details.

[1]BRIDGE, School of Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS, UK. [2]British Antarctic Survey, Geological Sciences Division, High Cross, Madingley Road, Cambridge CB3 0ET, UK. [3]Bristol Isotope Group, Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK. [4]School of Earth and Environment, Environment Building, University of Leeds, Leeds LS2 9JT, UK.

Figure 1 shows the GCM-predicted temperature and precipitation changes in the Greenland region associated with the four non-orbital hypotheses. The closure of the Panama seaway has a significant effect on the simulated global ocean circulation, leading to increased overturning in the Atlantic and increased northward transport of warm surface waters[20]. This ocean-driven temperature signal is advected over continental regions, and there is a resulting increase in summer temperature over Greenland (Fig. 1a). Associated with the surface warming is increased evaporation and precipitation. This strong correlation between temperature and precipitation change is common to all the simulations presented here. In particular, in this case, precipitation in the North Atlantic (where the changes in sea surface temperature are largest) increases significantly on closure of the Panama seaway. This results in increased snowfall over Greenland itself, in particular in the southeast (Fig. 1e), where the high-altitude coastal mountain ranges amplify the precipitation increase.

The termination of the permanent El Niño state leads to an increase in northward heat transport in the Atlantic, which increases surface temperatures in the North Atlantic, just south of Greenland (Fig. 1b). However, this temperature signal does not extend over Greenland itself, where the summer temperatures are relatively unchanged. There is an increase in annual mean precipitation over Greenland associated with a northward shift of the Atlantic storm track (Fig. 1f); again this change is concentrated in the southeast.

The uplift of the Rocky Mountains leads to local cooling in western North America because of the higher surface elevation and associated increase in snow cover, especially in winter. In summer, this signal is advected downstream over much of Canada and the North Atlantic, and into Greenland itself (Fig. 1c). Additionally, uplift on the west coast of Greenland leads to localized cooling. The cooler surface temperatures in the Northern Hemisphere result in a generally drier climate over Greenland following uplift; however, a slight northward deflection of the Atlantic storm track does lead to greater precipitation on the southern margins of Greenland (Fig. 1g).

Decreasing CO_2 during the Late Pliocene leads to global cooling of 1.3 °C. Consistent with simulations carried out under present-day boundary conditions[21], the temperature response is greatest at high latitudes, primarily owing to albedo feedbacks related to changes in snow and sea-ice coverage. In contrast, the cooler planet and increased meridional temperature gradient in the low-CO_2 simulation results in increased overturning strength in the North Atlantic, leading to greater northward oceanic heat transport, and a slight increase in temperature in the Barents Sea (Fig. 1d). Over Greenland, however, the direct CO_2 response dominates, and there is a summer cooling. The globally cooler climate results in generally decreased evaporation from the Earth's surface, and a global average decrease in precipitation. This is also true over Greenland itself, where the annual mean precipitation decreases (Fig. 1h), again consistent with high-CO_2 simulations carried out under present-day boundary conditions[21].

On first inspection the results shown in Fig. 1 could provide support for each of the hypotheses; for instance, both the Panama and ENSO anomalies are characterized by increases in precipitation over Greenland, and both the uplift and CO_2 anomalies are characterized by decreases in temperature. However, these climate anomalies are not sufficient on their own to draw conclusions about possible impacts on Greenland glaciation. The mass balance of an ice sheet is a complex function of the geographical and temporal distribution of accumulation, ablation and underlying bedrock, and is influenced by ice dynamics. Therefore we made use of a high-resolution ice-sheet model, which includes a representation of these processes, given the temperature and precipitation predictions from the GCM. The role of albedo feedbacks in these simulations is discussed in the Supplementary Information.

The resulting equilibrium configurations for the Greenland ice sheet are shown in Fig. 2. The closure of the Panama seaway actually reduces the volume of the ice sheet (compare the control configuration, Fig. 2c, with Fig. 2a). The warmer Greenland summer surface temperatures in the closed-seaway Pliocene control case result in increased ablation around the margins of the ice sheet, enough to remove the ice cap in the south of the island that is present in the open-seaway configuration. The increased precipitation results in a small increase in the maximum height of the ice cap in the east. The net change is a decrease of 0.8 m sea level equivalent (SLE) of the total ice volume on closure of the Panama seaway.

The termination of the permanent El Niño state also leads to a small decrease in volume of the ice sheet (0.3 m SLE; compare Fig. 2c with Fig. 2b). In this case, the limited response of the ice sheet is due to the relatively small changes in temperature and precipitation over Greenland (Fig. 1b, f).

Tectonic uplift does lead to a slight increase in Greenland ice volume (0.5 m SLE; compare Fig. 2d with Fig. 2c), with localized

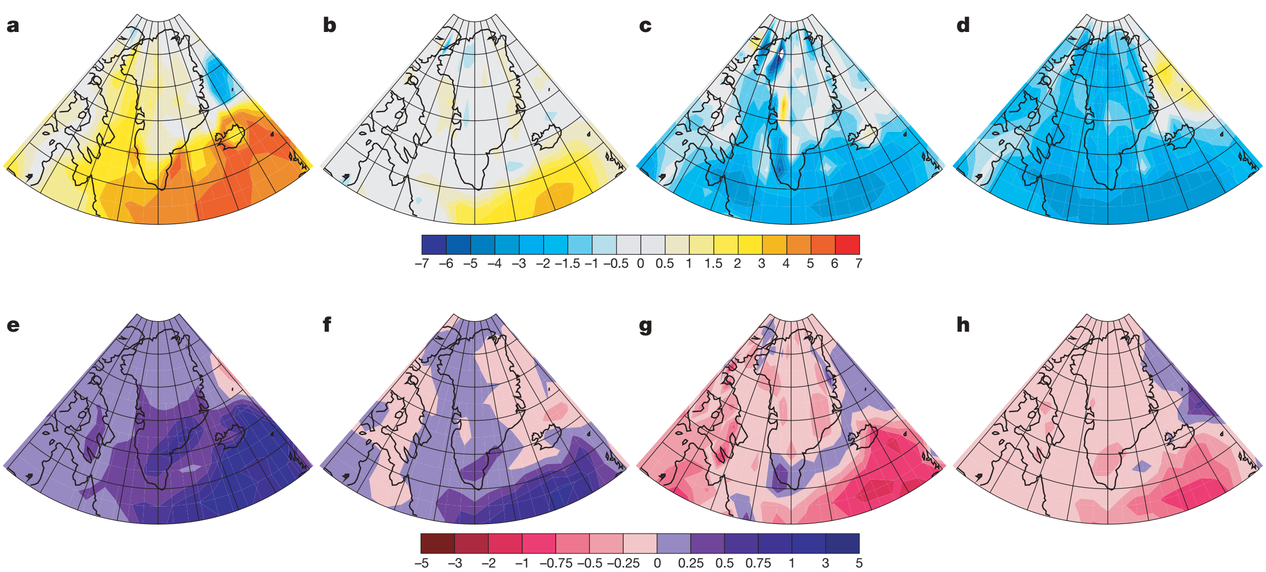


Figure 1 | Surface climate anomalies associated with the four hypotheses considered. a–d, Simulated surface temperature change (°C) during Northern Hemisphere summer due to closure of the Panama seaway (**a**), termination of a permanent El Niño state (**b**), tectonic uplift (**c**), decreasing CO_2 (**d**). **e–h,** Simulated annual precipitation change (mm day^{-1}) due to closure of the Panama seaway (**e**), termination of a permanent El Niño state (**f**), tectonic uplift (**g**), decreasing CO_2 (**h**).

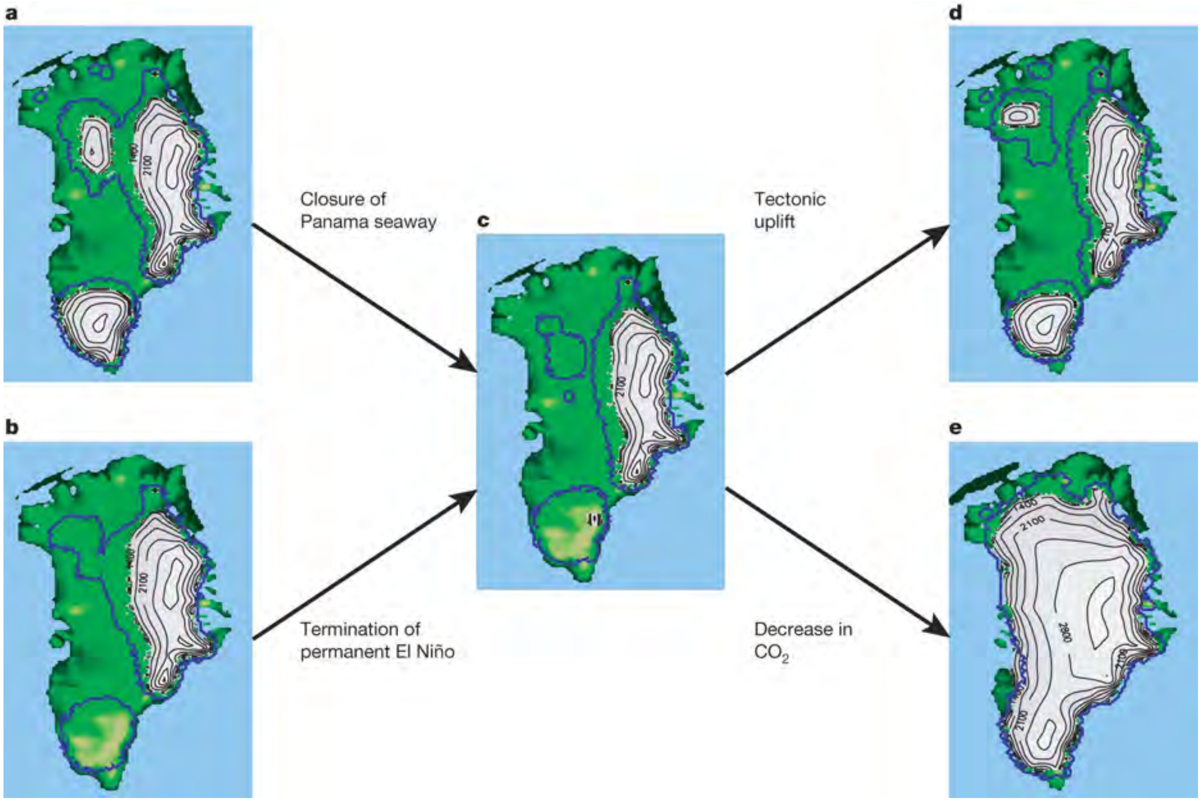


Figure 2 | Ice-sheet configurations. a–e, Pliocene ice-sheet configuration with open seaway (**a**), permanent El Niño (**b**), closed seaway, dynamic ENSO, low orography and high CO_2 (the Pliocene control) (**c**), high orography (**d**) and low CO_2 (**e**). The blue line represents the extent of the ice sheet under orbital conditions favourable for inception. The arrows represent the direction of geological time.

inception occurring in the south and northwest of the island, where temperature anomalies are the greatest (Fig. 1c). However, this is not extensive enough to be considered entry into full glaciation. Decreasing CO_2 has a considerable effect on the evolution of the ice sheet (compare Fig. 2e with Fig. 2c). The cooler summer temperatures under low-CO_2 conditions result in decreased ablation of the ice sheet, an effect that dominates over the decrease in accumulation. The net change is a large increase in volume of 6.3 m SLE. Notably, as well as being larger in volume, the low-CO_2 ice sheet also has the most extensive marine margin, capable of producing greater IRD.

To test the alternative hypothesis that one of these mechanisms acted in conjunction with favourable orbital conditions to produce an increase in glaciation, we repeated the five ice-sheet model simulations but included an additional time-varying forcing component, peaking at values appropriate for a cold orbit favourable for inception in the Northern Hemisphere. For more details see Methods and Supplementary Information.

The maximum ice-sheet extent from these transient orbital simulations, representing the Pliocene glacial configuration of the Greenland ice sheet, is represented in Fig. 2 as a blue line. As expected, the ice sheet expands for all five of these glacial simulations. However, with the exception of the larger low-CO_2 ice sheet, all of the ice sheets are similar in size. Therefore, there is no evidence that any of the tectonic and oceanographic changes modelled here have a 'priming' effect that promotes glaciation under favourable orbital conditions.

This study indicates that the decrease in atmospheric CO_2, from the high values of the mid-Pliocene to the lower values of the Quaternary, drove a significant increase in Greenland glaciation. It shows that the climatic shifts associated with the tectonically driven closure of the Panama seaway, with the termination of a permanent El Niño state, or with tectonic uplift, were not large enough to contribute significantly to this increase. In addition, it suggests that none of these processes acted as a priming mechanism for inception, triggered by orbital variations. It does, however, support the assertion that some ice did exist on Greenland before the onset of extensive glaciation, and that this waxed and waned on orbital timescales, a conclusion that is consistent with orbital timescale variations in benthic $^{18}O/^{16}O$ ratios before 3 Myr ago[22]. Further confidence in our results can be gained by noting that the dominant role of CO_2 compared with the other hypotheses is robust to variations in key parameters in the ice-sheet model, and also holds under warm-orbit conditions (see Supplementary Information). The work highlights the need for detailed multi-proxy CO_2 reconstructions through the Cenozoic, and expansion of the IRD record from the North Atlantic.

METHODS SUMMARY

We used the UK Met Office model HadCM3 (ref. 23) to carry out the suite of GCM simulations. As well as a Pliocene control, these were simulations with an open Panama seaway (to test the Panama hypothesis), a permanent El Niño (to test the ENSO hypothesis), modern orography (to test the uplift hypothesis), low-CO_2 (to test the CO_2 hypothesis) and a cold-orbit (to test the orbital-trigger hypothesis). We used the UK ice-sheet model GLIMMER[24] to carry out the ice-sheet model simulations. In the non-orbital ice-sheet simulations, the temperature and precipitation from the GCM were used to force the ice-sheet model to equilibrium. In the orbital ice-sheet simulations, the cold-orbit GCM was used to create a 10,000-year synthetic orbital cycle, covering half a precession cycle. For more details, and descriptions of the models used, see online Methods and Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 November 2007; accepted 27 June 2008.

1. Eldrett, J., Harding, I. C., Wilson, P. A., Butler, E. & Roberts, A. P. Continental ice in Greenland during the Eocene and Oligocene. *Nature* 446, 176–179 (2007).
2. Thiede, J. *et al.* Late Cenozoic history of the polar North Atlantic: Results from ocean drilling. *Quat. Sci. Rev.* 17, 185–208 (1998).
3. Maslin, M. A., Li, X. S., Loutre, M.-F. & Berger, A. The contribution of orbital forcing to the progressive intensification of Northern Hemisphere glaciation. *Quat. Sci. Rev.* 17, 411–426 (1998).

4. Bartoli, G. *et al.* Final closure of Panama and the onset of northern hemisphere glaciation. *Earth Planet. Sci. Lett.* **237,** 33–44 (2005).
5. Keigwin, L. D. Isotopic paleoceanography of the Caribbean and East Pacific: Role of Panama uplift in Late Neogene time. *Science* **217,** 350–353 (1982).
6. Haug, G. H. & Tiedemann, R. Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation. *Nature* **393,** 673–676 (1998).
7. Wara, M. W., Ravelo, A. C. & Delaney, M. L. Permanent El Niño-like conditions during the Pliocene warm period. *Science* **309,** 758–761 (2005).
8. Haywood, A. M., Dekens, P., Ravelo, A. C. & Williams, M. Warmer tropics during the mid-Pliocene? Evidence from alkenone paleothermometry and a fully coupled ocean–atmosphere GCM. *Geochem. Geophys. Geosyst.* **6,** doi:10.1029/2004GC000799 (2005).
9. Philander, S. G. & Fedorov, A. V. Role of tropics in changing the response to Milankovich forcing some three million years ago. *Paleoceanography* **18,** doi:10.1029/2002PA000837 (2003).
10. Ruddiman, W. F. & Kutzbach, J. E. Forcing of Late Cenozoic Northern Hemisphere climate by plateau uplift in southern Asia and the American West. *J. Geophys. Res.* **94,** 18409–18427 (1989).
11. Raymo, M. E. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate. *Nature* **359,** 117–122 (1992).
12. Greenwood, D. R. & Wing, S. L. Eocene continental climates and latitudinal temperature gradients. *Geology* **23,** 1044–1048 (1995).
13. Edgar, K. M., Wilson, P. A., Sexton, P. F. & Suganuma, Y. No extreme bipolar glaciation during the main Eocene calcite compensation shift. *Nature* **448,** 908–911 (2007).
14. Siegenthaler, U. *et al.* Stable carbon cycle-climate relationship during the Late Pleistocene. *Science* **310,** 1313–1317 (2005).
15. Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309,** 600–603 (2005).
16. Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406,** 695–699 (2000).
17. Pagani, M., Freeman, K. H. & Arthur, M. A. Late Miocene atmospheric CO_2 concentrations and the expansion of C_4 grasses. *Science* **285,** 876–879 (1999).
18. Raymo, M. E., Grant, B., Horowitz, M. & Rau, G. H. Mid-Pliocene warmth: stronger greenhouse and stronger conveyor. *Mar. Micropaleontol.* **27,** 313–326 (1996).
19. Kürschner, W. M., van der Burgh, J., Visscher, H. & Dilcher, D. L. Oak leaves as biosensors of late Neogene and early Pleiostocene paleoatmospheric CO_2 concentrations. *Mar. Micropaleontol.* **27,** 299–312 (1996).
20. Lunt, D. J., Valdes, P. J., Haywood, A. & Rutt, I. C. Closure of the Panama Seaway during the Pliocene: implications for climate and Northern Hemisphere glaciation. *Clim. Dyn.* **30,** 1–18 (2008).
21. Soloman, S. *et al. Climate Change 2007: The Physical Science Basis* (Contribution of Working Group I to the Fourth, Assessment Report of the IPCC, Cambridge Univ. Press, 2007).
22. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* **20,** doi:10.1029/2004PA001071 (2005).
23. Gordon, C. *et al.* The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* **16,** 147–168 (2000).
24. Payne, A. J. A thermomehanical model of ice flow in West Antarctica. *Clim. Dyn.* **15,** 115–125 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was carried out in the framework of the British Antarctic Survey GEACEP (Greenhouse to ice-house: Evolution of the Antarctic Cryosphere and Palaeoenvironment) programme. D.J.L. is funded by BAS and RCUK fellowships. G.L.F. is funded by a NERC research fellowship. E.J.S. is funded by a NERC studentship.

Author Contributions D.J.L. carried out the GCM and ice-sheet model simulations, except for the permanent El Niño GCM simulation, which was provided by P. Valdes at the University of Bristol, UK. A.M.H. contributed to setting up the Pliocene control and permanent El Niño GCM simulations. D.J.L., G.L.F. and A.M.H. were involved in the study design. E.J.S. devised parts of the ice-sheet model driver and experimental set-up. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.J.L. (d.j.lunt@bristol.ac.uk).

METHODS

Here we describe the GCM and ice-sheet simulations analysed in this paper. Further details can be found in the Supplementary Information.

The Pliocene control GCM simulation is a 580-year extension of the equivalent Pliocene control simulation described by Haywood and Valdes[25]. The open-seaway simulation is a 200-year extension of the equivalent simulation of Lunt *et al.*[20]. This allows a closer degree of equilibrium to be obtained than in those studies. The permanent El Niño simulation is the same as the 'PlioPacific^{Dateline}' simulation discussed by Haywood *et al.*[26]. The uplift and CO_2 simulations are both 200 years long, and are initialized from the end of the Pliocene control simulation of Lunt *et al.*[20].

The Pliocene control GCM simulation conforms to the PRISM2 (Pliocene Research, Interpretation and Synoptic Mapping-2) standard[25]. It has been found to be in reasonable agreement with global proxy data and, in particular, in excellent agreement with Arctic palaeobotanical data and palaeotemperature estimates[27].

To test the Panama hypothesis we carried out an open-seaway perturbation simulation, representing a time period prior to the control, in which a connection of 350 m depth was made between the Pacific and Atlantic in the region that forms the Panama Isthmus in the Pliocene control simulation, in agreement with proxy data appropriate for this time interval[28].

To test the ENSO hypothesis, we carried out a simulation in which Pacific sea surface temperatures between 40° north and south of the Equator were forced towards a constant value, typical of the western Pacific. This represents the permanent El Niño state which may have existed earlier than the control, and is in contrast to the fully dynamic ENSO of the Pliocene control simulation.

To test the importance of orographic changes, we carried out a high-orography perturbation simulation, representing a time period after the control, in which the orography was prescribed to be that of the present day, in contrast to the PRISM2 orography used in the Pliocene control simulation. The exception is in the glaciated regions of Greenland and Antarctica, where the Pliocene orography was used. The most important changes are in North America, where the Rocky Mountains in the PRISM2 data set are up to a kilometre lower than the present day. There are also regions of low orography in the PRISM2 data set on the west and east coasts of Greenland itself, which is consistent with data indicating late Cenozoic uplift around the North Atlantic[29].

To test the CO_2 hypothesis, we carried out a low-CO_2 simulation, again representing a time period after the control. In this simulation, atmospheric CO_2 was prescribed to be 280 p.p.m.v., a value appropriate for the pre-industrial period, in contrast to 400 p.p.m.v. in the Pliocene control, which is at the upper end of mid-Pliocene atmospheric CO_2 estimates[18] from carbon isotope ($\delta^{13}C$) data. To test the importance of orbital effects, we carried out a Pliocene cold-orbit simulation, in which the orbital configuration was set to be identical to that of 115,000 years before present (115 kyr BP, the most recent glacial inception, following the last interglacial). The orbit at 115 kyr BP is characterized by cold boreal summers, and at 65° north in July results in less insolation than any orbital configuration between 5 and 2 Myr ago[30] (see Supplementary Information).

More details of the experimental set-up, spin-up and boundary conditions of all the GCM simulations are given in the Supplementary Information.

For the non-orbital ice-sheet simulations, we used the temperature and precipitation fields from the GCM to force the ice-sheet model for a total of 50,000 years, long enough for equilibrium with the driving climate to be reached. The GCM simulations were carried out using the modern orbital configuration, which results in an insolation at 65° north in July close to, but slightly colder than, the long-term average of the past 5 million years. The non-orbital ice-sheet simulations are therefore intended to represent a long-term average configuration of the Greenland ice sheet, in equilibrium with near-average orbital conditions. More details of the experimental set-up of the ice-sheet simulations are given in the Supplementary Information.

The transient orbital ice-sheet experiments represent the forcing experienced during a transition from near-average orbital conditions to glacial orbital conditions, and back again. The orbital-trigger hypothesis states that under favourable orbital conditions, one of the Panama, uplift, ENSO or CO_2 mechanisms could have produced a significant change in volume of the Greenland ice sheet. However, given that the orbital forcing is transient, it is not appropriate just to repeat the five GCM simulations including a constant favourable orbit. This would overestimate the orbital effect, as in reality the insolation at high latitudes varies in response to the ~20-kyr precessional cycle, the 'favourable' orbit being the maximum of this cycle.

Therefore, we create a synthetic orbital cycle, by applying a time-varying anomaly on top of the five original GCM simulations, and use this to force the ice-sheet model. We simulate the 'cold' half of the cycle (10 kyr) only, as it is this that is appropriate for testing the orbital-trigger hypothesis. The maximum of this anomaly (calculated relative to the Pliocene control with modern orbit) is obtained from the GCM simulation carried out under orbital conditions for 115 kyr BP, but otherwise Pliocene boundary conditions. For each orbital ice-sheet simulation, the initial condition is the equilibrium state from the corresponding non-orbital simulation. This corresponds to the fact that the non-orbital experiments are simulating an average ice-sheet configuration over many orbital cycles, whereas the orbital experiments represent a deviation from this long-term average state. Given that the 115-kyr orbit is fairly extreme in the context of the past 5 million years, this provides a strong test of the orbital-trigger hypothesis—for example, if the control cold-orbit ice sheet were significantly larger than the open-seaway cold-orbit ice sheet, this would imply that the closure of the Panama seaway acted as a 'priming' mechanism for glaciation, triggered by the favourable orbital conditions. More details of the orbital ice-sheet simulations are given in the Supplementary Information.

25. Haywood, A. M. & Valdes, P. J. Modelling Middle Pliocene warmth: contribution of atmosphere, oceans and cryosphere. *Earth Planet. Sci. Lett.* **218,** 363–377 (2004).
26. Haywood, A. M., Valdes, P. J. & Peck, V. L. A permanent El Niño-like state during the Pliocene? *Paleoceanography* **22,** doi:10.1029/2006PA001323 (2007).
27. Hill, D. J., Haywood, A. M., Hindmarsh, R. C. A. & Valdes, P. J. in *Deep Time Perspectives on Climate Change: Marrying the Signal from Computer Models and Biological Proxies* (eds Williams, M., Haywood, A. M., Gregory, F. J. & Schmidt, D. N.) *Micropalaeontol. Soc. Spec. Publ.* **2,** 517–538 (Geol. Soc. London, 2007).
28. Schmidt, D. N. in *Deep Time Perspectives on Climate Change: Marrying the Signal from Computer Models and Biological Proxies* (eds Williams, M., Haywood, A. M., Gregory, F. J. & Schmidt, D. N.) *Micropalaeontol. Soc. Spec. Publ.* **2,** 427–442 (Geol. Soc. London, 2007).
29. Japsen, P. & Chalmers, J. A. Neogene uplift and tectonics around the North Atlantic: overview. *Global Planet. Change* **24,** 165–173 (2000).
30. Berger, A. & Loutre, M.-F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10,** 297–317 (1991).

LETTERS

Formation of current coils in geodynamo simulations

Akira Kageyama[1], Takehiro Miyagoshi[1] & Tetsuya Sato[1]

Computer simulations have been playing an important role in the development of our understanding of the geodynamo[1–3], but direct numerical simulation of the geodynamo with a realistic parameter regime is still beyond the power of today's supercomputers. Difficulties in simulating the geodynamo arise from the extreme conditions of the core, which are characterized by very large or very small values of the non-dimensional parameters of the system. Among them, the Ekman number, *E*, has been adopted as a barometer of the distance of simulations from real core conditions, in which *E* is of the order of 10^{-15}. Following the initial computer simulations of the geodynamo[4,5], the Ekman number achieved has been steadily decreasing, with recent geodynamo simulations[6–8] performed with *E* of the order of 10^{-6}. Here we present a geodynamo simulation with an Ekman number of the order of 10^{-7}—the highest-resolution simulation yet achieved, making use of 4,096 processors of the Earth Simulator. We have found that both the convection flow and magnetic field structures are qualitatively different from those found in larger-Ekman-number dynamos. The convection takes the form of sheet plumes or radial sheet jets[9], rather than the columnar cell structures[10] that are usually found. We have found that this sheet plume convection is an effective dynamo and the generated current is organized as a set of coils in the shape of helical springs or at times as a torus.

The Earth's outer core is liquid iron in convective motion. To model the dynamo process in the outer core, we have performed a three-dimensional numerical simulation of a magnetohydrodynamic (MHD) dynamo in a rotating spherical shell: an electrically conducting fluid is confined between two concentric and co-rotating spheres with inner and outer radii of $r_i = 0.3$ and $r_o = 1.0$, respectively. The two spheres rotate with the same angular velocity Ω. The temperatures of the inner and outer spheres are kept hot and cold, respectively. Gravity works towards the centre of the spheres. We apply weak perturbations to the temperature and magnetic fields of the initial (unstable) condition. Thermal convection sets in, and the flow generates the magnetic field through the so-called MHD dynamo process. The Rayleigh number, Ra (ref. 11), measured at the bottom of the shell is 1.5×10^{10}, which is 300–1,000 times larger than the critical value for the onset of convection. Both the Prandtl and magnetic Prandtl numbers are unity. The Ekman number[3] defined by $E = \nu / 2\Omega r_o^2$ is 2.3×10^{-7} in this simulation, where ν is viscosity. We believe that this is the lowest Ekman number achieved to date in geodynamo simulations.

The smaller the Ekman number, the more difficult it is to perform the simulation, as it requires higher resolution and a higher parallelization rate of the code[12]. By making use of a newly developed spherical grid system, the 'Yin-Yang' grid[13,14], we have achieved a high resolution geodynamo simulation for a full spherical shell region. The Yin-Yang grid is a kind of overset grid[15], applied to the spherical geometry. Two identical grids, the Yin grid and the Yang grid, are combined with partial overlap to cover the full spherical shell region. The grid size is 511 (in r) $\times$ 514 (in θ) $\times$ 1,538 (in ϕ) $\times$ 2 (Yin and Yang), with r radius ($0.3 \le r \le 1.0$), θ co-latitude ($\pi/4 \le \theta \le 3\pi/4$), and ϕ longitude ($-3\pi/4 \le \phi \le 3\pi/4$). For this simulation, we have used 512 nodes or 4,096 processors of the Earth Simulator, which is the maximum size allowed for a calculation.

It is broadly accepted that the convection flow in a rapidly rotating spherical shell may be described as a set of (time dependent) columnar convection cells[16]. The convection columns are straight and parallel to the spherical rotation axis, due to the strong constraint of the Coriolis force (Taylor-Proudman's theorem). However, in our low-*E* regime (of the order of 10^{-7}), we have found that the convection is made of a rather different basic structure—fine-scale thin jet sheets or sheet plumes.

In the growing phase of convection in our simulation, the flow is formed in multicellular columns piled in the radial direction[17], but after saturation, the convection takes a rather different form. Figure 1a shows the z or axial component of the vorticity, ω_z, in the equatorial and meridional planes at time $t = 430$ in our simulation unit, which is normalized by the sound wave crossing time of r_o. The major flow is composed of many plumes elongated in the radial direction, rather than columns. Analysing the velocity vectors in detail, we have found that the plume structure is composed of jet flow in the positive s-direction and in the negative s-direction, side by side; here we use cylindrical coordinates (s, ϕ, z) for description only. Supplementary Fig. 1 shows the velocity field in the equatorial plane by vector arrows. The width w of the jet plumes is very thin ($w \approx 0.025$), staying almost constant as radius s increases[18]. The azimuthal Fourier mode number, m, of the flow is about 50 at $s = 0.4$. The diffusion time τ_d scaled by w is $\tau_d = w^2/\eta = 34$ (where η is resistivity), which is much shorter than our calculation time of 430.

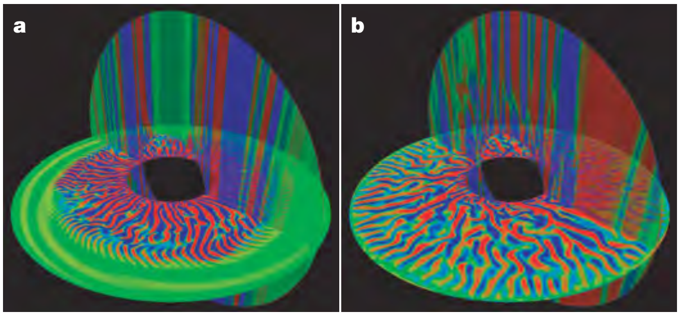


Figure 1 | Equatorial and meridional cross-sections of the axial component of the vorticity, ω_z. The Ekman number, *E*, is 2.3×10^{-7} in **a** and 2.6×10^{-6} in **b**. Convection plumes are evident in the equatorial cross-sections. The meridional cross-sections show that the flow is nearly two-dimensional. The convection in these low-Ekman-number regimes is organized as a set of thin plume sheets, rather than columnar cells. It can be seen that the larger Ekman number in **b** leads to thicker plume sheets.

[1]Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology, Yokohama, 236-0001, Japan.

As the meridional cross-section (Fig. 1a) shows, z-dependence of the velocity is very weak: the convection can be described as a set of sheet-like plumes. These sheets are not stationary; the timescale of the pattern changes may be estimated (roughly) from the sheet width w and the flow velocity v (timescale $\approx w/v$). The typical magnetic Reynolds number $\mathrm{Rm} = dv/\eta = 700$, where d is shell depth ($d = r_o - r_i$). The Elsasser number Λ is of the order of 1–10. For a comparison, we show simulation results for a larger Ekman number, $E = 2.6 \times 10^{-6}$ (Fig. 1b); this figure suggests that a larger Ekman number leads to thicker plume sheets.

Sheet plume convection has already been discovered in laboratory convection experiments on water in a hemispherical shell by Sumita and Olson[9]. Figure 2a shows a snapshot of a convection pattern of their experiment visualized by flakes, seen through the transparent equatorial plane. The Ekman number in their experiment was $E = 1 \times 10^{-6}$. Figure 2b shows the distribution of ω_z of our simulation visualized by volume rendering viewed from the north. As the flakes are sensitive to velocity shear, the visualization of ω_z in Fig. 2b is a natural counterpart of Fig. 2a. Note that the sheet plumes branch off, going in an outward (positive s) direction.

Another important characteristic of the flow observed in our simulation is the existence of zonal flow near the outer boundary. The direction of the zonal flow is westward. Figure 1a shows an indistinct spiral pattern with four arms, which is embedded in the zonal flow with azimuthal mode number $m = 0$. The power of the zonal flow, measured at the radius in the equator, is an order of magnitude larger than other modes. The co-existence of the sheet plumes and the zonal flow is a new feature that was not observed in the experiments by Sumita and Olson[9]. This zonal flow region near the outer spherical boundary is absent when $E = 2.6 \times 10^{-6}$ (Fig. 1b). This could be explained by a weak constraint imposed by the Coriolis force. We are analysing the formation mechanism of the zonal flow, and it will be reported in the near future.

By observing the plume pattern as a movie in the equatorial plane, we found that the plumes can be divided into two parts—a lower part (near the inner core) and an upper part (near and above the middle depth of the spherical shell). The lower part drifts eastward, and upper part drifts westward. We note that the convection velocity develops faster than the magnetic field. The sheet plume structure is formed before the magnetic energy becomes large enough to affect the velocity field. Therefore, the sheet structure is not formed or maintained by the magnetic field. This is in contrast with a dynamo simulation under a low-E regime in a box geometry[19], in which small scale flows are strongly affected by the Lorentz force.

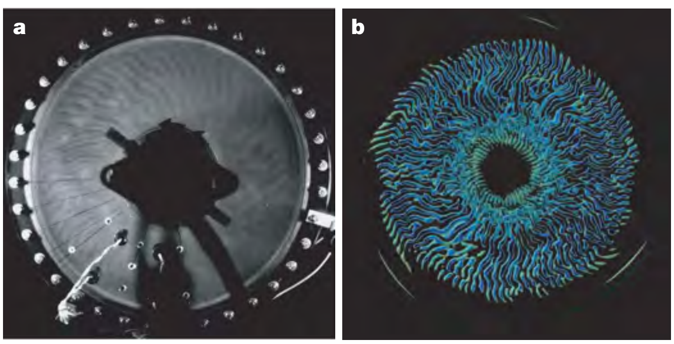


Figure 2 | Sheet plumes observed in experiment and simulation. **a**, The experiment by Sumita and Olson[9]. **b**, Our simulation. In Sumita and Olson's experiment, a rotating hemispherical shell is filled with water. The centrifugal force plus gravity works in the radial direction. The cooled inner sphere causes thermal convection. This is a snapshot, visualized by flakes seen through the transparent equatorial plane; the Ekman number $E = 1 \times 10^{-6}$ in this experiment. Similar sheet plumes can be observed in our simulation (**b**; $E = 2.3 \times 10^{-7}$). This is a visualization by volume rendering of the axial vorticity.

The sheet plume convection is an effective dynamo. The total magnetic energy at $t = 430$ (Fig. 1a) is about 4 times larger than the total convection energy. The typical Alfvén velocity is 0.04 in our units. The magnetic field is generated mainly at mid-latitude and outside the tangential cylinder. This is confirmed by examining the distribution of the dynamo source term $D \equiv -\mathbf{v}\cdot(\mathbf{j} \times \mathbf{B})$ in the spherical shell, where $\mathbf{v}$, $\mathbf{j}$ and $\mathbf{B}$ are respectively velocity, current and magnetic field. By analysing distributions of D, $\mathbf{v}$ and $\mathbf{B}$, we have found that there are two kinds of mechanism by which the magnetic field can be generated in this simulation; in the first kind, the magnetic energy is generated by downward ($-v_s$) plume flows. In this case, typical $\mathbf{B}$ is perpendicular to $\mathbf{v}$ and the field lines are drawn by the downward $\mathbf{v}$ against the field tension force. This dynamo mechanism is essentially the same as observed in our laminar dynamo simulation[20]. The second kind of dynamo mechanism is very different: in this case, $\mathbf{B}$ is parallel to $\mathbf{v}$. The magnetic field is generated by upward ($v_s > 0$) plume flows. The flow is accelerating in the positive s direction after a plume leaves from the inner hot boundary. The magnetic field lines embedded in the plume are stretched in the parallel direction by this acceleration. From mass conservation, stretching of a magnetic flux tube causes reduction of the tube cross-section, and therefore, flux density amplification. As the total volume of the flux tube is nearly constant, the total magnetic energy is increased by this stretching process. This magnetic field amplification by field line stretching is described as a classical picture of the MHD dynamo mechanism[21], and it has recently observed in geodynamo simulations[22].

Stretching by the parallel flow makes the magnetic field both strong and straight. Thus a straight flux tube is formed that is embedded in the upward-moving sheet plume. The localized straight flux tube leads to the formation of a current coil around the tube. Figure 3 shows a whole current field structure visualized by current lines, started from a sphere at $r = 0.46$ with nearly uniform distribution of starting points. We can see that there are many current coils. The distribution of the current coils is relatively localized in the north and south mid-latitude. By analysing the current field in detail by virtual reality visualization[23,24], we have found that most of the current coils are in the form of a helical spring, but that some of them are in a torus configuration. Plotting the distribution of the dynamo source term D, we have confirmed that the current coils are located in regions of strong positive D. The magnetic field is generated at the current coils. Figure 4 shows typical current coils with magnetic field lines; a helical spring at higher latitude and a torus at lower latitude. As Fig. 3 suggests, the axial dipole field is not dominant in this simulation. Almost all the spherical harmonic modes with degrees from $l = 1$ (dipole) to $l \approx 10$ are of the same order of magnitude. In this sense, this could be regarded as a kind of general dynamo simulation, rather than a reproduction of the geodynamo.

To analyse the magnetic field of our simulation data, we first drew (many) magnetic field lines, but the resulting pictures were too complicated to understand. However, by drawing current lines, the hidden structure of the magnetic field was disclosed. We found that the current coil structure survives, though in a less distinctive fashion, in the higher-Ekman-number ($E = 2.6 \times 10^{-6}$) case; see Supplementary Fig. 2.

Our simulation indicates that the convection and its dynamo enters a new regime when we decrease the Ekman number to the order of 10^{-7}. Dynamo scalings[7,25] established by simulations with $E \geq 10^{-6}$ should be checked in this new dynamo regime. The current coils and straight magnetic flux tubes found in this simulation may be relevant to the solar dynamo, as sunspots are formed by emerging magnetic flux tubes. Finally, we note that the magnetic Prandtl number ($\mathrm{Pm} = 1$) in this simulation is far from the Earth's outer-core value, which is of the order of 10^{-6}—the next challenge is to find the effects of different values of Pm.

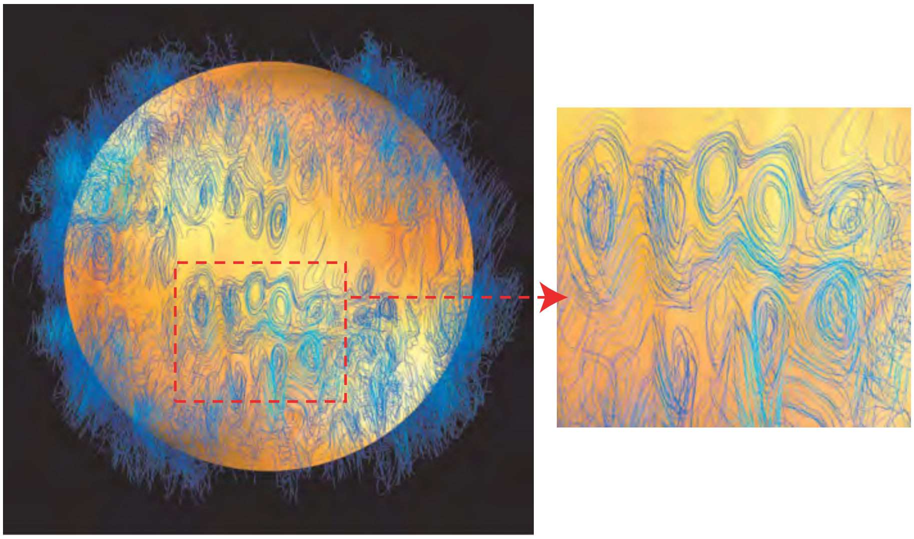


Figure 3 | Current field structure near the inner core at $t = 430$. This is the same simulation as shown in Fig. 1a. The blue lines are current field lines. The left image shows the distribution in the spherical shell. The tracings of the current lines start from a sphere at $r = 0.46$. The orange sphere of $r = 0.41$ is placed to help the view. The current field is organized as a set of 'coils', located mainly in the middle latitudes. The right panel is a magnified image. Most of the current coils are in the form of a helical spring, but some of them form a torus.

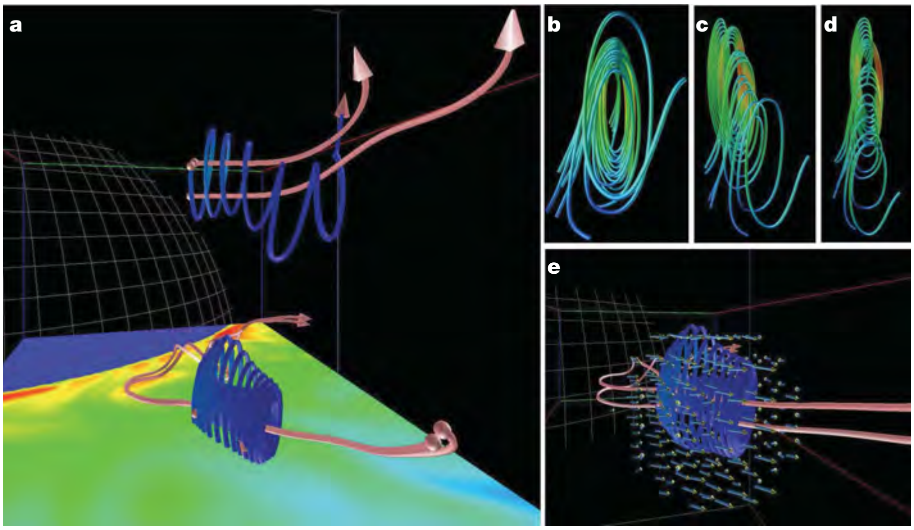


Figure 4 | Detailed structure of two typical current coils (blue) with their associated magnetic field lines (pink). There are two helical coils in **a**. The upper one is a helical spring and the lower one is a torus. The coloured plane in **a** is a cross-section perpendicular to the z-axis in the northern hemisphere. (This is not the equatorial plane.) The colour in the plane denotes the mass density; the fluid with high density (red) is sinking ($v_s < 0$), while the fluid with low density (blue) is rising. Panels **b**, **c** and **d** show the helical spring coil of **a** (the upper blue coil) viewed from different angles. This is a typical helical spring shape of the current coils. Panel **e** shows the current torus of **a** (that is, the lower blue coil), with velocity fields (arrows). The velocity fields at the field lines are parallel to the magnetic field. This flow, which is a part of the sheet plume, is accelerating along the field line and it stretches out to form a straight magnetic flux tube. The magnetic field lines are located in the outward $v_s > 0$ sheet plume. **a** and **e** are snapshots taken by the virtual reality visualization software VFIVE[23,24] in the CAVE system[29].

METHODS SUMMARY

The physical model, basic equation, and boundary conditions adopted in this simulation are the same as used in our previous work[4,20,26–28]. The compressible MHD equations in the spherical shell geometry are discretized by the second-order finite difference method. The magnetic field has only a radial component on the inner and outer spherical boundaries. The time integration method is fourth-order Runge-Kutta. To avoid grid convergence near the poles of a spherical coordinate system, we have made use of the Yin-Yang grid system[13,14] in this simulation. We have used virtual reality visualization software VFIVE[23,24] in the CAVE virtual reality system[29] to obtain Fig. 4a and e. The current torus cell structure was first found by this virtual reality visualization.

Received 2 May; accepted 27 June 2008.

1. Roberts, P. H. & Glatzmaier, G. A. Geodynamo theory and simulations. *Rev. Mod. Phys.* **72,** 1081–1123 (2000).
2. Jones, C. A. Convection-driven geodynamo models. *Phil. Trans. R. Soc. Lond. A* **358,** 873–897 (2000).
3. Kono, M. & Roberts, P. H. Recent geodynamo simulations and observations of the geomagnetic field. *Rev. Geophys.* **40,** 1013, doi:10.1029/2000RG000102 (2002).
4. Kageyama, A. *et al.* Computer simulation of a magnetohydrodynamic dynamo. II. *Phys. Plasmas* **2,** 1421–1431 (1995).
5. Glatzmaier, G. A. & Roberts, P. H. A three-dimensional self-consistent computer simulation of a geomagnetic field reversal. *Nature* **377,** 203–209 (1995).
6. Takahashi, F., Matsushima, M. & Honkura, Y. Simulations of a quasi-Taylor state geomagnetic field including polarity reversals on the Earth simulator. *Science* **309,** 459–461 (2005).
7. Christensen, U. & Aubert, J. Scaling properties of convection-driven dynamos in rotating spherical shells and applications to planetary magnetic fields. *Geophys. J. Int.* **166,** 97–114 (2006).
8. Sakuraba, A. & Hamano, Y. Turbulent structure in Earth's fluid core inferred from time series of geomagnetic dipole moment. *Geophys. Res. Lett.* **34,** L15308, doi:10.1029/2007GL029898 (2007).

9. Sumita, I. & Olson, P. Laboratory experiments on high Rayleigh number thermal convection in a rapidly rotating hemispherical shell. *Phys. Earth Planet. Inter.* **117**, 153–170 (2000).
10. Busse, F. A model of the geodynamo. *Geophys. J. R. Astron. Soc.* **42**, 437–459 (1975).
11. Gilman, P. & Glatzmaier, G. Compressible convection in a rotating spherical shell. I. Anelastic equations. *Astrophys. J.* **45** (Suppl.), 335–349 (1981).
12. Zhang, K. & Schubert, G. Comparison of terrestrial and solar dynamos. *Rep. Prog. Phys.* **69**, 1581–1605 (2006).
13. Kageyama, A. & Sato, T. ''Yin-Yang grid'': An overset grid in spherical geometry. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2004GC000734 (2004).
14. Kageyama, A. *et al.* A 15.2 TFlops simulation of geodynamo on the earth simulator. *Proc. ACM/IEEE SC2004* Conf.–35–43 (2004).
15. Chesshire, G. & Henshaw, W. Composite overlapping meshes for the solution of partial differential equations. *J. Comput. Phys.* **90**, 1–64 (1990).
16. Busse, F. Convective flows in rapidly rotating spheres and their dynamo action. *Phys. Fluids* **14**, 1301–1314 (2002).
17. Plaut, E. & Busse, F. Multicellular convection in rotating annuli. *J. Fluid Mech.* **528**, 119–133 (2005).
18. Aubert, J., Brito, D., Nataf, H.-C., Cardin, P. & Masson, J.-P. A systematic experimental study of rapidly rotating spherical convection in water and liquid gallium. *Phys. Earth Planet. Inter.* **128**, 51–74 (2001).
19. Stellmach, S. & Hansen, U. Cartesian convection driven dynamos at low Ekman number. *Phys. Rev. E* **70**, 056312 (2004).
20. Kageyama, A. & Sato, T. Generation mechanism of a dipole field by a magnetohydrodynamic dynamo. *Phys. Rev. E* **55**, 4617–4626 (1997).
21. Davidson, P. *An Introduction to Magnetohydrodynamics* (Cambridge Univ. Press, 2001).
22. Aubert, J., Aurnou, J. & Wicht, J. The magnetic structure of convection-driven numerical dynamos. *Geophys. J. Int.* **172**, 945–956 (2008).
23. Kageyama, A., Tamura, Y. & Sato, T. Visualization of vector field by virtual reality. *Prog. Theor. Phys.* **138** (Suppl.), 665–673 (2000).
24. Ohno, N. & Kageyama, A. Scientific visualization of geophysical simulation data by the CAVE VR system with volume rendering. *Phys. Earth Planet. Inter.* **163**, 305–311 (2007).
25. Olson, P. & Christensen, U. R. Dipole moment scaling for convection-driven planetary dynamos. *Earth Planet. Sci. Lett.* **250**, 561–571 (2006).
26. Ochi, M. M., Kageyama, A. & Sato, T. Dipole and octupole field reversals in a rotating spherical shell: Magnetohydrodynamic dynamo simulation. *Phys. Plasmas* **6**, 777–787 (1999).
27. Kageyama, A., Ochi, M. M. & Sato, T. Flip-flop transitions of the magnetic intensity and polarity reversals in the magnetohydrodynamic dynamo. *Phys. Rev. Lett.* **82**, 5409–5412 (1999).
28. Li, J., Sato, T. & Kageyama, A. Repeated and sudden reversals of the dipole field generated by a spherical dynamo action. *Science* **295**, 1887–1890 (2002).
29. Cruz-Neira, C., Sandin, D. & DeFanti, T. Surrounded-screen projection-based virtual reality. *Proc. SIGGRAPH* **93**, 135–142 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Olson and J. Aurnou for pointing out the close connection between the sheet plume structure observed in this simulation and the experiments of ref. 9. We thank I. Sumita for a detailed explanation of his laboratory experiments and comments, and for providing Fig. 2a. We thank N. Ohno for helping with the visualization. This work was supported by KAKENHI (17540404) and The Mitsubishi Foundation.

Author Contributions A.K. was involved in project planning, simulation code development, simulation runs, data analysis and manuscript preparation. T.M. was involved in a part of the simulation runs, data analysis and manuscript preparation. T.S. was involved in project planning and manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.K. (kage@jamstec.go.jp).

LETTERS

Acetylcholine contributes through muscarinic receptors to attentional modulation in V1

J. L. Herrero[1]*, M. J. Roberts[1]*†, L. S. Delicato[1]*†, M. A. Gieselmann[1], P. Dayan[2] & A. Thiele[1]

Attention exerts a strong influence over neuronal processing in cortical areas[1,2]. It selectively increases firing rates[2–4] and affects tuning properties[1,5], including changing receptive field locations and sizes[3,6]. Although these effects are well studied, their cellular mechanisms are poorly understood. To study the cellular mechanisms, we combined iontophoretic pharmacological analysis of cholinergic receptors with single cell recordings in V1 while rhesus macaque monkeys (*Macaca mulatta*) performed a task that demanded top-down spatial attention. Attending to the receptive field of the V1 neuron under study caused an increase in firing rates. Here we show that this attentional modulation was enhanced by low doses of acetylcholine. Furthermore, applying the muscarinic antagonist scopolamine reduced attentional modulation, whereas the nicotinic antagonist mecamylamine had no systematic effect. These results demonstrate that muscarinic cholinergic mechanisms play a central part in mediating the effects of attention in V1.

Attention is a rich and complex psychological and neurobiological construct. In both non-human[1–4] and human primates[7], researchers have investigated a wealth of paradigms tapping visual attention, showing the selection of behaviourally relevant over behaviourally irrelevant stimuli, and the enhancement of the former's processing. Further modulation of sensory areas is assumed to be driven by the frontal and parietal cortex[7,8] through direct cortico–cortical feedback connections. However, frontal regions also influence sensory areas indirectly, through connections to cholinergic neurons in the basal forebrain that have ascending projections to sensory areas[9,10], and there is ample evidence for the involvement of acetylcholine (ACh) in attentional modulation[10–14]. For instance, depletion of ACh in cortical areas by disrupting cholinergic fibres originating in the basal forebrain results in persistent attentional impairments[10,15], and attentional deficits seen in Alzheimer's disease partly arise from cholinergic dysfunction[16]. The precise nature of the contribution of ACh to attentional modulation in the cortex is at present unclear. Thus, we recorded the strength of attentional modulation in neurons from V1 of three macaque monkeys, while simultaneously performing pharmacological analysis of cholinergic receptor contributions. Subjects performed a task demanding voluntary allocation of attention[5,17] under control conditions and when ACh, or muscarinic or nicotinic receptor antagonists, were iontophoretically applied in the vicinity of the neurons under study[5,17].

We used identical task, training, and surgical and neurophysiological procedures as previously described[5,17] (see Methods Summary and Supplementary Methods). Neurons were activated by bar stimuli of optimal orientations centred on their receptive fields. Attention was manipulated by a visual cue, which guided attention to be in or away from the neurons' receptive fields (the attend-receptive-field and attend-away conditions, respectively). The effect of attention on the firing rates of neurons was tested in 15–25 trials in each condition: control, drug-applied and recovery (see Supplementary Methods).

We recorded 156 neurons from 3 monkeys (26 neurons in monkey B, 87 in monkey HU and 43 in monkey HO) in the absence and presence of ACh application (Supplementary Methods provides more details). We tested whether attention, bar length or drug application significantly affected neuronal activity, and whether there was a significant interaction between these factors (three-factor analysis of variance (ANOVA), $P < 0.05$). We used the response period from 200 ms to 500 ms after stimulus onset for our analysis, as attentional modulation is most prominent during the sustained response in the striate[4,5] and extrastriate cortex[18].

Neurons were analysed further if attention and drug application significantly affected firing rates, or if a significant interaction between attention and drug application occurred. Strengths of attentional modulation were quantified by calculating the area under the receiver operating characteristic (ROC[5]) curve on the basis of single-trial responses (given knowledge of the bar length). ROC values of 0.5 indicate that an ideal observer can only perform at chance level in deciding the locus of attention. Higher ROC values indicate greater attentional response enhancement, thus if ACh increased attentional modulation, ROC values should be increased in ACh-applied conditions relative to control conditions.

A total of 86 out of 156 neurons (16 in monkey B, 38 from monkey HU and 32 from monkey HO) passed the basic statistical test (three-factor ANOVA) on application of ACh. As seen in Fig. 1a, ACh generally increased attentional modulation. The neuron showed a significant main effect of attention ($F_{1,284} = 24.92$, $P < 0.001$) and a significant effect of drug ($F_{1,284} = 6.7$, $P = 0.01$). ROC values showed that attentional modulation was increased on ACh application (ROC (−ACh): 0.6774 (0.8° bar length); ROC (+ACh): 0.8262; for further example cells, see Supplementary Methods). The population ROC values for the drug and control condition are shown in Fig. 2a. In the presence of ACh, ROC values were increased. A two-factor ANOVA showed that the effect was significant for the population of cells and did not depend on bar length (data were converted to *z*-scores before testing; samples passed normality and equal variance tests ($P > 0.05$); factor 1, drug application: $F_{1,874} = 16.58$, $P < 0.001$ (two-factor ANOVA); factor 2, bar length: $F_{6,874} = 7.61$, $P < 0.001$, drug × bar lengths interaction: $F_{6,874} = 1.058$, $P = 0.386$). The effect of ACh application on attentional modulation was significant in monkey HO and monkey HU individually ($P < 0.05$), but not in monkey B. However, the trend in monkey B was the same as that in the other two monkeys, and the lack of significance was possibly due to the smaller neuronal sample. Figure 2b shows the time course of the effect of ACh on attentional modulation. The upper and lower

[1]Institute of Neuroscience, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. [2]Gatsby Computational Neuroscience Unit, University College London, 17 Queens Square, London WCIN 3AR, UK. †Present addresses: F. C. Donders Centre for Cognitive Neuroimaging, PO Box 9101, 6500 HB Nijmegen, The Netherlands (M.J.R.); Department of Psychology, School of Business, Law & Psychology, University of Sunderland, Sunderland SR6 0DD, UK (L.S.D.).
*These authors contributed equally to this work.

green curves show normalized population activities when ACh was applied for attention into and away from the receptive fields, respectively. The upper and lower red curves show the same, but without ACh. Thus, the widths of the colour-shaded regions show the evolution of attentional modulation. The fact that the green areas are generally wider than the red ones is an indication of the boosting effect of ACh on attentional modulation.

ACh slightly increased the overall firing rates (Fig. 2b), making it possible that increased ROCs were simply caused by multiplicative scaling of firing rates. However, it can be shown that ROC values are invariant to proportional rescaling of distributions (see Supplementary Methods). Furthermore, we calculated a modulation index for each cell, drug condition and bar length: modulation index = (activity$_{\text{attend receptive field}}$ − activity$_{\text{attend away}}$)/(activity$_{\text{attend receptive field}}$ + activity$_{\text{attend away}}$) which normalizes explicitly for firing rate. In line with the ROC analysis, we found that ACh significantly increased the population modulation index ($F_{1,874} = 9.49$, $P = 0.002$, two-factor ANOVA), and that this was independent of bar length ($F_{6,874} = 0.42$, $P = 0.869$).

Having established that ACh application increases attentional modulation, we were interested in the receptors that mediate this effect. Thus we recorded 118 cells (46 from monkey B and 72 from monkey HU) while monkeys performed the same task, but we applied the muscarinic antagonist scopolamine iontophoretically (see Supplementary Methods). Out of the 118 cells, 41 showed a significant attention and a significant drug effect, or a significant interaction between these two (20 cells from monkey B and 21 cells from monkey HU).

Scopolamine generally reduced attentional modulation. An example cell is shown in Fig. 1b. Attention ($F_{1,464} = 33.34$, $P < 0.001$), drug application ($F_{1,464} = 18.13$, $P < 0.001$) and bar length ($F_{3,464} = 3.13$, $P = 0.025$) all had a significant effect on firing rates (three-factor ANOVA). Attentional modulation was significantly reduced in the presence of scopolamine (drug × attention interaction: $F_{3,464} = 13.5$, $P < 0.001$). We quantified the effect of scopolamine application on attentional modulation by calculating the ROC for each cell and bar length. Population ROCs are shown in Fig. 3a. Scopolamine significantly reduced attentional modulation ($F_{1,466} = 27.744$, $P < 0.001$, two-factor ANOVA). The effect of scopolamine on ROCs was independent of bar length (drug × bar length interaction: $F_{6,466} = 0.217$, $P = 0.971$). The reduction in attentional modulation with scopolamine application was significant for each monkey individually ($P < 0.001$).

Figure 3b shows that the effect of scopolamine on the evolution of the population response in the same form as in Fig. 2b. The green shaded regions (scopolamine application) are now thinner than the red regions during the sustained response, indicating that attentional modulation was stronger in the absence of scopolamine application. Scopolamine also reduced firing rates, so, as for ACh, we also calculated the firing-rate-normalized modulation index. Confirming the results from the ROC analysis, modulation indexes were significantly reduced on scopolamine application ($F_{1,466} = 11.15$, $P = 0.0009$; see Supplementary Methods).

ACh also acts on nicotinic receptors, which might equally contribute to attentional modulation in V1. We therefore tested attentional modulation in the presence and absence of the nicotinic antagonist mecamylamine, whilst recording 151 V1 cells (113 cells from monkey HU and 39 cells from monkey HO). Significant effects of attention and of mecamylamine application, or an interaction, were found in 65 cells (47 cells from monkey HU and 18 cells from monkey HO). Figure 4 shows the mean ROC values for these cells as a function of bar length for the control and the mecamylamine-applied condition. A two-factor ANOVA did not show significant effects of drug

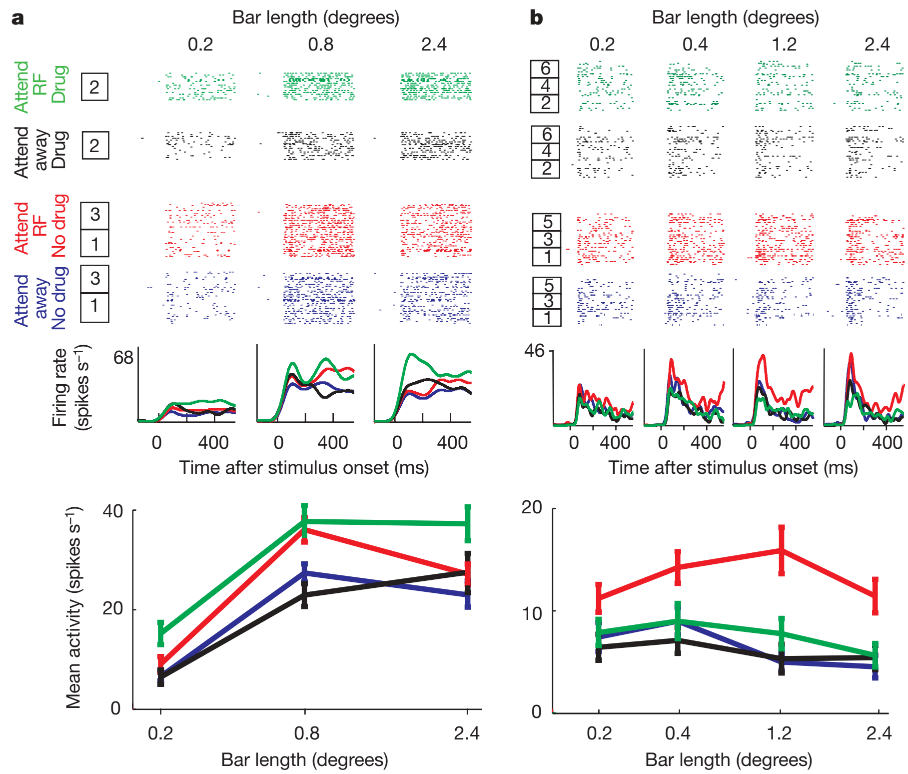


Figure 1 | Drug effects on attentional modulation. a, Attentional enhancement by application of ACh. We recorded the effect of attention on firing rates when no ACh was applied (16 trials) for three bar lengths in the attend-inside receptive field (RF) versus attend-outside (away) condition (box 1). Thereafter we recorded the effect of attention when ACh was applied (16 trials; box 2) followed by recovery (16 trials; box 3). The bottom plot shows the average activity (from 200 ms to 500 ms after stimulus onset) for the different stimulus, attention and drug conditions. ACh increased attentional modulation. **b**, Effect of scopolamine application on neuronal attentional modulation. All conventions are as in **a**. Drug application and recovery were repeated three times (boxes 1–6). Scopolamine reduced attentional modulation. Error bars represent s.e.m.

application ($F_{1,744} = 0.54$, $P = 0.465$), whereas there was a significant effect of bar length ($F_{6,744} = 2.98$, $P = 0.0069$) on the size of the ROC values. There was no interaction between mecamylamine applied/not-applied and bar length ($F_{6,744} = 0.40$, $P = 0.881$). These findings were consistent for the two animals. The same outcome was true when the modulation index was used to quantify the effects of mecamylamine application on attentional modulation ($F_{1,744} = 1.41$, $P = 0.236$, two-factor ANOVA).

None of the effects of ACh, scopolamine or mecamylamine were due to differences in eye position induced by drug application (see Supplementary Methods).

For the neuromodulator dopamine it has been reported that only an optimal dose benefits neuronal and behavioural performance[19], whereby too much and too little dopamine is detrimental for cognitive performance. We found evidence for a similar pattern of results for ACh in V1: that is, only small doses of extra ACh increased attentional modulation (for details, see Supplementary Methods).

Systemic applications of cholinergic drugs or large injections into higher cortical areas have been shown to affect performance in humans[20] and macaque monkeys[21]. The highly local character of iontophoretic drug application used in our study made it unlikely that performance would be affected, as even systemic infusions result in relatively small changes to reaction times[20]. In line with this expectation, neither scopolamine nor mecamylamine application affected reaction times significantly. However, when ACh was applied we found a significant interaction between drug application and the locus of attention on reaction times ($F_{1,31827} = 4.56$, $P = 0.038$, three-factor ANOVA). ACh increased reaction times in the attend-away condition and decreased reaction times in the attend-receptive-field condition. This effect was significant across the three monkeys, but only individually in monkey HO ($F_{1,5588} = 4.77$, $P = 0.029$, three-factor ANOVA). Similar trends were found in monkey HU and monkey B, but they did not reach significance (interaction between ACh application and locus of attention on reaction times: monkey HU, $F_{1,20334} = 3.03$, $P = 0.081$; monkey B, $F_{1,5905} = 3.06$, $P = 0.080$). These results indicate that ACh boosts aspects of attention when the animals attend to the stimulus in the receptive field, thus decreasing reaction times. The increase in reaction times in the attend-away condition could be due to improved neuronal processing at the application site making it more difficult to disengage attention from the irrelevant location, thus slightly increasing reaction times (for more information, see Supplementary Methods).

We found that ACh contributes to attentional modulation in V1, of the macaque monkey through muscarinic, but not nicotinic, receptor mechanisms. Fewer excitatory cells compared to inhibitory cells are subject to muscarinic modulation in V1 (ref. 22), raising the possibility that our recordings were mostly from inhibitory neurons.

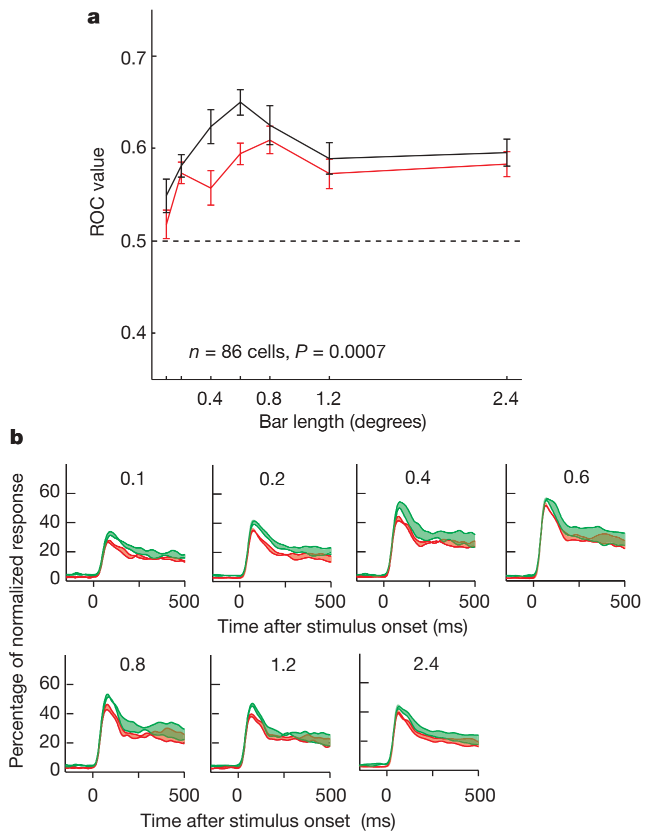


Figure 2 | Acetylcholine effects on attentional modulation. **a**, Quantification of attentional modulation by mean population ROC for different bar length with (black line) or without (red line) ACh application (86 cells; error bars denote s.e.m.). **b**, Normalized population response depending on ACh application, stimulus (bar length; indicated at the top of each subplot) and attention condition. Activity levels were normalized relative to the peak activity of each neuron and averaged across the population. The red lines indicate activity without ACh application; green lines represent activity with ACh application. The upper line of each colour plot shows the normalized activity when attention was directed to the neuron's receptive field, the lower line when it was directed away from the receptive field. Widths of the coloured areas show strength of attentional modulation. ACh increased attentional modulation.

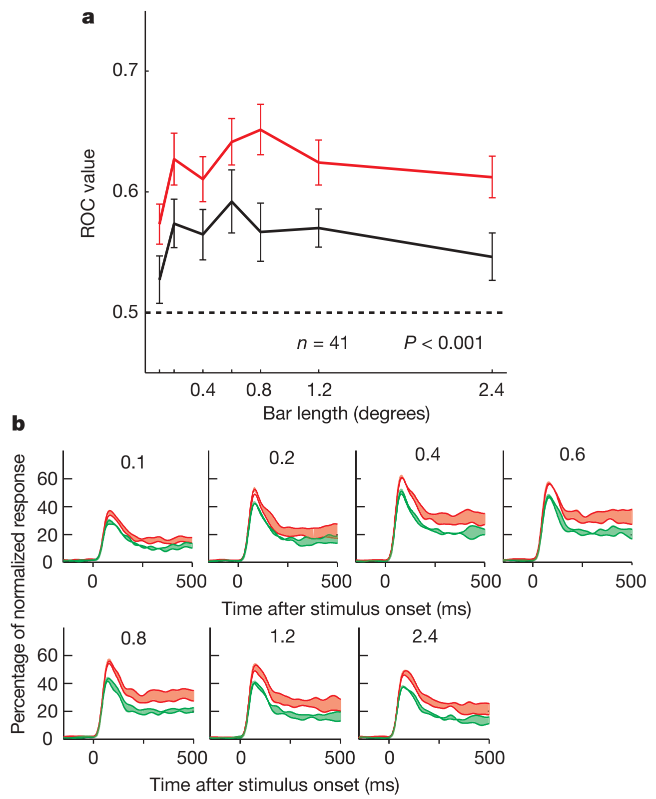


Figure 3 | Effect of the muscarinic antagonist scopolamine on attentional modulation. **a**, Quantification of attentional modulation by mean population ROC when scopolamine was (black line) and was not (red line) applied ($n = 41$ cells, error bars represent s.e.m.). ROC values were significantly reduced on scopolamine application, demonstrating that muscarinic receptors are involved in mediating the effects of attention in V1. **b**, Normalized population response depending on scopolamine application, bar length (indicated at the top of each subplot) and attention. The widths of the red and green shaded areas show the strengths of attentional modulation as a function of scopolamine application (green lines represent activity with scopolamine application; red lines denote activity without scopolamine application). Attentional modulation was reduced when scopolamine was applied.

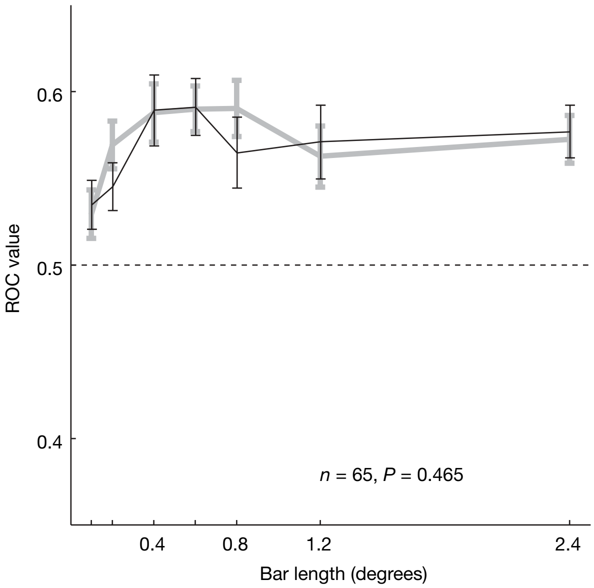


Figure 4 | Effect of the nicotinic antagonist mecamylamine on attentional modulation. Attentional modulation was quantified by calculating the population ROC for the different bar lengths when mecamylamine was (black line) and was not (grey line) applied ($n = 65$ cells). ROC values were not significantly affected by mecamylamine application. Error bars show the s.e.m.

Analysis of the 'spike' waveform, however, suggests that recordings were mostly from excitatory neurons. Thus, the effects may have been mediated through local cortical circuitry, perhaps through an effect on the sort of oscillatory activity that has been linked to attentional effects[23]. Alternatively, the fact that attentional effects are stronger in V2 and V4 (ref. 24), where there are also more muscarinic receptors[22] on excitatory cells, suggests that the effects may have been more direct and not necessarily mediated by changes in the overall network state.

Nicotinic receptors in macaque V1 are mostly located presynaptically on thalamocortical terminals in layer 4C, where they affect the neuronal gain[25]. Although attention affects firing rates in simple cells of V1 (ref. 26), where nicotinic receptors would be expected to have the most direct effect[25], our result indicates that this is not mediated at the thalamocortical synaptic input stage. Although we found no evidence of nicotinic receptor contribution to attentional modulation in V1 it is probable that they contribute to attentional modulation in higher areas[27], perhaps through the feed-forward cortico–cortical pathway.

An important question for neuromodulatory accounts of attention[28] is reconciling the highly localized and fast effects of spatial attention with the apparent (though perhaps arguable[29]) coarseness of the ACh innervation. One possibility is that an interaction between ACh and glutamatergic feedback mediates attentional processing in sensory areas. Increased amounts of ACh alter the strengths of the connections in V1 and the biophysical state of sensory neurons which may then allow spatially specific glutamatergic feedback to enhance specific incoming information. Other possible sources of highly localized cholinergic contribution to attentional modulation are the recently described cholinergic cortical interneurons[30].

Future studies will be necessary to clarify these issues. Nevertheless, the first step in understanding attentional modulation is to unpick its cortical determinants; our findings directly address this.

METHODS SUMMARY

We recorded extracellular neuronal activity from three hemispheres in three male macaque monkeys (*Macaca mulatta*), whilst applying ACh, scopolamine or mecamylamine iontophoretically on selected trials. Animals were implanted with a custom made head-holding device and recording chambers made of Tecapeek GF for compatibility in functional magnetic resonance imaging settings. Surgical procedures were performed under aseptic conditions and general anaesthesia. Experiments and surgeries were performed in accordance with the European Communities Council Directive 1986 (86/609/EEC), the National Institutes of Health (Guidelines for Care and Use of Animals for Experimental Procedures), the Society for Neurosciences Policies on the Use of Animals and Humans in Neuroscience Research, and the UK Animals Scientific Procedures Act.

The monkey's task was to detect a small change in luminance at a cued (attended) location, while ignoring a change that occurred at a non-cued location and fixating a central fixation spot throughout the trial. After a fixation-only period, two identical stimuli were presented (test stimuli): one centred on the receptive field and the other at the same eccentricity in the opposite hemi-field. After 500–800 ms (randomized in 1 ms steps) a patch appeared at the centre of one of the bars. If presented in the cued location the monkey had to release the touch bar within 500 ms to receive a juice reward. If presented in the un-cued location the monkey had to continue to hold the touch bar and maintain fixation until target appearance. This occurred 1,000–1,300 ms (randomized in 1 ms steps) after the distracter appeared. Thus, we recorded activity when animals attended to the receptive field of the neuron under study and when they attended away from it. We then compared activity levels for these attentional conditions with and without drug application.

For all further information about the paradigm, the neuronal recordings, iontophoresis and data analysis, see Supplementary Methods.

Received 4 April; accepted 6 June 2008.
Published online 16 July 2008.

1. Spitzer, H., Desimone, R. & Moran, J. Increased attention enhances both behavioral and neuronal performance. *Science* **240**, 338–340 (1988).
2. Treue, S. & Maunsell, J. H. R. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539–541 (1996).
3. Reynolds, J. H., Chelazzi, L. & Desimone, R. Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* **19**, 1736–1753 (1999).
4. Roelfsema, P. R., Lamme, V. A. & Spekreijse, H. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* **395**, 376–381 (1998).
5. Roberts, M., Delicato, L. S., Herrero, J., Gieselmann, M. A. & Thiele, A. Attention alters spatial integration in macaque V1 in an eccentricity-dependent manner. *Nature Neurosci.* **10**, 1483–1491 (2007).
6. Womelsdorf, T., Anton-Erxleben, K., Pieper, F. & Treue, S. Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nature Neurosci.* **9**, 1156–1160 (2006).
7. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nature Rev. Neurosci.* **3**, 201–215 (2002).
8. Moore, T. & Armstrong, K. M. Selective gating of visual signals by microstimulation of frontal cortex. *Nature* **421**, 370–373 (2003).
9. Russchen, F. T., Amaral, D. G. & Price, J. L. The afferent connections of the substantia innominata in the monkey, *Macaca fascicularis*. *J. Comp. Neurol.* **242**, 1–27 (1985).
10. Sarter, M., Hasselmo, M. E., Bruno, J. P. & Givens, B. Unraveling the attentional functions of cortical cholinergic inputs: interactions between signal-driven and cognitive modulation of signal detection. *Brain Res. Rev.* **48**, 98–111 (2005).
11. Furey, M. L., Pietrini, P., Haxby, J. V. & Drevets, W. C. Selective effects of cholinergic modulation on task performance during selective attention. *Neuropsychopharmacology* **33**, 913–923 (2008).
12. Robbins, T. W. Chemistry of the mind: neurochemical modulation of prefrontal cortical function. *J. Comp. Neurol.* **493**, 140–146 (2005).
13. Parikh, V., Kozak, R., Martinez, V. & Sarter, M. Prefrontal acetylcholine release controls cue detection on multiple timescales. *Neuron* **56**, 141–154 (2007).
14. Nelson, C. L., Sarter, M. & Bruno, J. P. Prefrontal cortical modulation of acetylcholine release in posterior parietal cortex. *Neuroscience* **132**, 347–359 (2005).
15. McGaughy, J., Dalley, J. W., Morrison, C. H., Everitt, B. J. & Robbins, T. W. Selective behavioral and neurochemical effects of cholinergic lesions produced by intrabasalis infusions of 192 IgG-saporin on attentional performance in a five-choice serial reaction time task. *J. Neurosci.* **22**, 1905–1913 (2002).
16. Nobili, L. & Sannita, W. G. Cholinergic modulation, visual function and Alzheimer's dementia. *Vision Res.* **37**, 3559–3571 (1997).
17. Thiele, A., Delicato, L. S., Roberts, M. J. & Gieselmann, M. A. A novel electrode-pipette design for simultaneous recording of extracellular spikes and iontophoretic drug application in awake behaving monkeys. *J. Neurosci. Methods* **158**, 207–211 (2006).
18. Reynolds, J. H., Pasternak, T. & Desimone, R. Attention increases sensitivity of V4 neurons. *Neuron* **26**, 703–714 (2000).
19. Goldman-Rakic, P. S., Muly, E. C. III & Williams, G. V. D_1 receptors in prefrontal cells and circuits. *Brain Res. Rev.* **31**, 295–301 (2000).
20. Witte, E. A., Davidson, M. C. & Marrocco, R. T. Effects of altering brain cholinergic activity on covert orienting of attention: comparison of monkey and human performance. *Psychopharmacology (Berl.)* **132**, 324–334 (1997).
21. Davidson, M. C. & Marrocco, R. T. Local infusion of scopolamine into intraparietal cortex slows covert orienting in rhesus monkeys. *J. Neurophysiol.* **83**, 1536–1549 (2000).

22. Disney, A. A., Domakonda, K. V. & Aoki, C. Differential expression of muscarinic acetylcholine receptors across excitatory and inhibitory cells in visual cortical areas V1 and V2 of the macaque monkey. *J. Comp. Neurol.* **499,** 49–63 (2006).
23. Fries, P., Reynolds, J. H., Rorie, A. E. & Desimone, R. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* **291,** 1560–1563 (2001).
24. Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77,** 24–42 (1997).
25. Disney, A. A., Aoki, C. & Hawken, M. J. Gain modulation by nicotine in macaque V1. *Neuron* **56,** 701–713 (2007).
26. McAdams, C. J. & Reid, R. C. Attention modulates the responses of simple cells in monkey primary visual cortex. *J. Neurosci.* **25,** 11023–11033 (2005).
27. Thiel, C. M., Zilles, K. & Fink, G. R. Nicotine modulates reorienting of visuospatial attention and neural activity in human parietal cortex. *Neuropsychopharmacology* **30,** 810–820 (2005).
28. Yu, A. J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46,** 681–692 (2005).
29. Fournier, G. N., Semba, K. & Rasmusson, D. D. Modality- and region-specific acetylcholine release in the rat neocortex. *Neuroscience* **126,** 257–262 (2004).
30. von Engelhardt, J., Eliava, M., Meyer, A. H., Rozov, A. & Monyer, H. Functional characterization of intrinsic cholinergic interneurons in the cortex. *J. Neurosci.* **27,** 5633–5642 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The work was supported by the BBSRC (BBS/B/09325), the Wellcome Trust (070380/Z/03/Z) and the Gatsby Charitable Foundation.

Author Contributions J.L.H., M.J.R., L.S.D. and M.A.G. performed the experimental work; A.T. and P.D. did the project planning; A.T., J.L.H. and M.J.R. did the data analysis; A.T. and P.D. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.T. (alex.thiele@ncl.ac.uk).

Vol 454 | 28 August 2008 | doi:10.1038/nature07168

A blend of small molecules regulates both mating and development in *Caenorhabditis elegans*

Jagan Srinivasan[1]*, Fatma Kaplan[2,3,4]*, Ramadan Ajredini[2,3,4], Cherian Zachariah[2,3,4], Hans T. Alborn[5], Peter E. A. Teal[5], Rabia U. Malik[6], Arthur S. Edison[2,3,4], Paul W. Sternberg[1] & Frank C. Schroeder[6]

In many organisms, population-density sensing and sexual attraction rely on small-molecule-based signalling systems[1,2]. In the nematode *Caenorhabditis elegans*, population density is monitored through specific glycosides of the dideoxysugar ascarylose (the 'ascarosides') that promote entry into an alternative larval stage, the non-feeding and highly persistent dauer stage[3,4]. In addition, adult *C. elegans* males are attracted to hermaphrodites by a previously unidentified small-molecule signal[5,6]. Here we show, by means of combinatorial activity-guided fractionation of the *C. elegans* metabolome, that the mating signal consists of a synergistic blend of three dauer-inducing ascarosides, which we call ascr#2, ascr#3 and ascr#4. This blend of ascarosides acts as a potent male attractant at very low concentrations, whereas at the higher concentrations required for dauer formation the compounds no longer attract males and instead deter hermaphrodites. The ascarosides ascr#2 and ascr#3 carry different, but overlapping, information, as ascr#3 is more potent as a male attractant than ascr#2, whereas ascr#2 is slightly more potent than ascr#3 in promoting dauer formation[7]. We demonstrate that ascr#2, ascr#3 and ascr#4 are strongly synergistic, and that two types of neuron, the amphid single-ciliated sensory neuron type K (ASK) and the male-specific cephalic companion neuron (CEM), are required for male attraction by ascr#3. On the basis of these results, male attraction and dauer formation in *C. elegans* appear as alternative behavioural responses to a common set of signalling molecules. The ascaroside signalling system thus connects reproductive and developmental pathways and represents a unique example of structure- and concentration-dependent differential activity of signalling molecules.

C. elegans hermaphrodites produce a chemical signal that strongly attracts males[5,6]. To identify this signal, we developed a new protocol for obtaining samples of secreted metabolites from different life stages of *C. elegans* (Supplementary Information). Biological activity of these samples was confirmed using a bioassay in which time spent by males or hermaphrodites within the vicinity of a chemical sample was measured (Fig. 1a, Supplementary Movie 1, Supplementary Methods). To determine the timing of mating pheromone release, samples from all life stages (egg, L1, dauer, L2, L3, L4, young adult and adult) were tested for biological activity on both males and hermaphrodites (Fig. 1b). *C. elegans* males were strongly attracted to L4, young adult and adult samples, whereas hermaphrodites were not attracted. Consistent with the observation that *C. elegans* reaches sexual maturity at the end of the L4 stage[8] (Fig. 1b), these results indicated that L4, young adult and adult hermaphrodites secrete a chemical signal that specifically attracts males.

To characterize the mating signal, we subjected samples derived from young adults to a multistep fractionation scheme, starting with C_{18}-reverse-phase solid-phase extraction chromatography. Strong male attraction was observed for one of the resulting fractions (Fig. 1c), which was further fractionated using coupled ion-exchange columns. Of the resulting seven fractions A to G, none was active at physiologically relevant concentrations when tested individually (data not shown), which suggested that male attraction depends on the synergy of two or more signalling molecules. To determine which fractions were required for activity, we assayed a series of combinations of fractions A–G; these assays showed that combination of fractions F or G with fraction A produced significant activity (Fig. 1d).

Because fraction A appeared to be required for full activity, we characterized it using nuclear magnetic resonance (NMR) spectroscopy[9] and liquid chromatography-mass spectrometry (LC-MS) (Supplementary Figs 2 and 8 and Supplementary Table 1). Two-dimensional NMR spectroscopic analyses suggested that the major component of fraction A is a novel derivative of 5-*O*-ascarylosyl-5*R*-hydroxy-2-hexanone, or ascr#2 (according to proposed nomenclature for nematode compounds) (Fig. 2a, Supplementary Figs 3 and 4), which was recently shown to induce dauer formation in *C. elegans*[7]. Additional NMR spectroscopic analyses showed that the major component of fraction A features a β-glucosyl substituent attached to C2 of the ascarylose in ascr#2 (Supplementary Figs 5–7). These assignments were corroborated by LC-MS analyses that showed an *m*/*z* value of 426 to represent the ammonium adduct ($M + NH_4^+$) of a compound with a nominal mass of 408 AMU and molecular formula $C_{18}H_{32}O_{10}$. Comparison of these spectroscopic data with those of synthetic sample of 5-*O*-(2′-*O*-[β-D-glucosyl]-ascarylosyl)-5*R*-hydroxy-2-hexanone provided final proof for the identify of the major component of fraction A, which we named ascr#4 (Fig. 2a).

The identification of ascr#4 in fraction A suggested that ascarosides might have a role as mating signals. Therefore, we analysed the fractions required for activity, A, F and G, for the presence of additional ascarosides. LC-MS analyses revealed the presence of ascr#3 in fractions F and G, as well as of small amounts of ascr#2 in fraction A (Supplementary Fig. 8). Next we tested synthetic samples of the three identified ascarosides, ascr#2, ascr#3 and ascr#4, for activity, using the assay described in Fig. 1a (Fig. 2a). Consistent with the assay results for fractions A, F and G, none of the three compounds was active at physiological concentrations when tested individually (data not shown). However, at higher concentrations, ascr#2 and ascr#3 were both active (Fig. 2b). The corresponding dose–response curves show a strongly biphasic activity profile (Fig. 2b), which is characteristic of many types of pheromone[10]. In contrast to ascr#2 and ascr#3, ascr#4 alone was not active in the male attraction assay, even at

[1]Howard Hughes Medical Institute and Biology Division, California Institute of Technology, 1200 E. California Boulevard, Pasadena, California 91125, USA. [2]Department of Biochemistry and Molecular Biology, [3]McKnight Brain Institute, [4]National High Magnetic Field Laboratory, University of Florida, PO Box 100245, Gainesville, Florida 32610-0245, USA. [5]Center for Medical, Agricultural and Veterinary Entomology, USDA-ARS, 1600–1700 SW 23rd Drive, PO Box 14565, Gainesville, Florida 32604, USA. [6]Boyce Thompson Institute, Cornell University, Ithaca, New York 14853, USA.
*These authors contributed equally to this work.

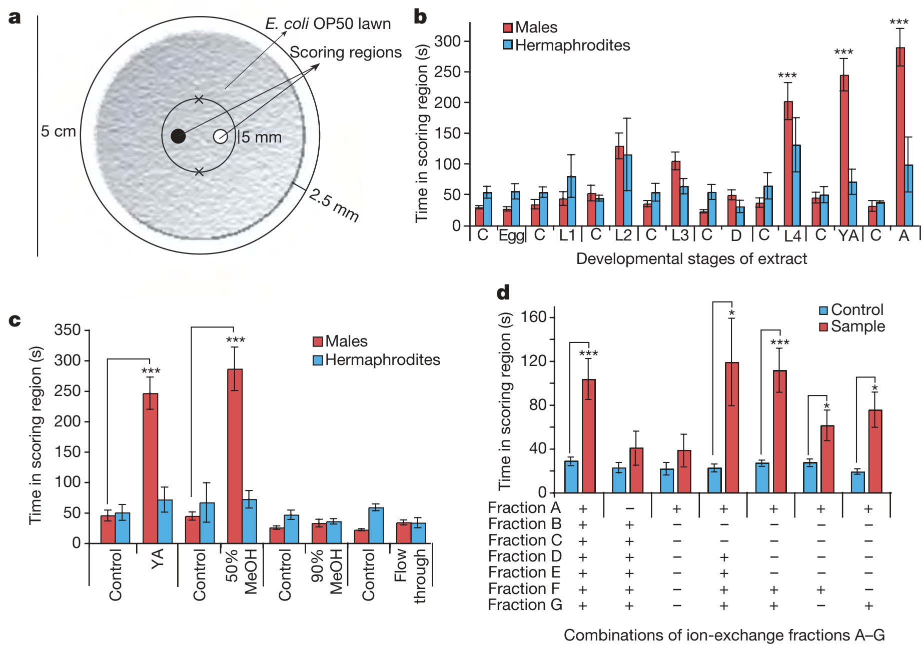


Figure 1 | Activity-guided fractionation of worm metabolites. **a**, Representation of the bioassay used to measure mating behaviour in worms. Crosses mark the initial positions of the assayed animals (see Supplementary Methods). **b**, Male and hermaphrodite responses to secreted metabolites produced by hermaphrodites at different developmental stages. L1–L4, the first four larval stages; D, dauer stage; YA, young adult stage; A, adult stage; C, control. $n \geq 30$ animals for each histogram. **c**, Assay results for C_{18}-reversed-phase chromatography fractions of young adult metabolite extract. **d**, Assay results for combinations of ion-exchange fractions of the active fraction from **c**: A, neutral; B, 250 mM KCl anion; C, 250 mM KCl cation; D, 500 mM KCl anion; E, 500 mM KCl cation; F, 1,000 mM KCl cation; G, 1,000 mM KCl anion. Error bars, s.e.m.; $*P < 0.01$, $***P < 0.0001$, unpaired t-test (see Supplementary Methods).

concentrations much greater than physiological levels. Because the assay results for the ion-exchange fractions (Fig. 1d) suggested that the mating signal consists of multiple compounds that act synergistically, we combined ascr#2 and ascr#3 in amounts that did not elicit significant male attraction when assayed individually (10 fmol ascr#3 and 100 fmol ascr#2; arrows in Fig. 2b). This mix produced strong male attraction, demonstrating synergism of ascr#2 and ascr#3 (Fig. 2c).

At the concentrations of ascr#2 and ascr#3 found in fractions A and F, a mixture of these two compounds also produced significant activity, but was less potent than the combination of fractions A and F (Fig. 2d), suggesting synergy with a third component, perhaps ascr#4.

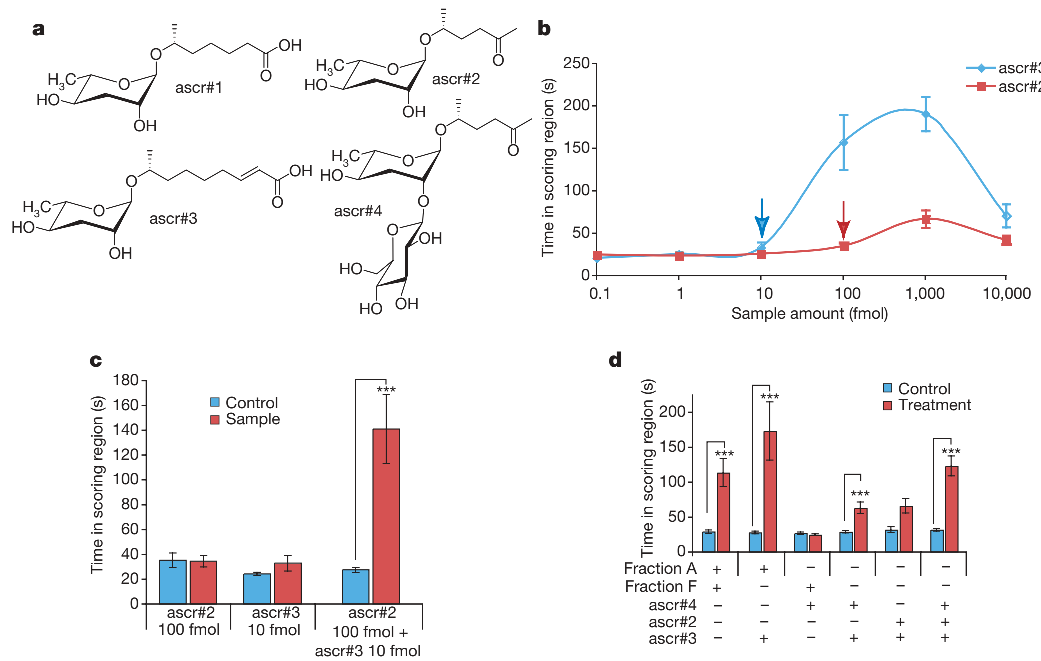


Figure 2 | Synergy between ascr#2, ascr#3 and ascr#4. **a**, Structures of ascr#1, ascr#2, ascr#3 and ascr#4. **b**, Dose–response curves of ascr#2 and ascr#3. **c**, Synergistic effects of ascr#2 and ascr#3 from points respectively indicated by red and blue arrows in **b**. **d**, Demonstration that the three synthetic compounds account for most of the mating activity. For all entries, ascr#4 was tested at 1 pmol and ascr#2 and ascr#3 were each tested at 20 fmol. $n \geq 30$ animals for each histogram. Error bars, s.e.m.; $*P < 0.01$, $***P < 0.0001$, unpaired t-test.

Additional tests using ternary mixtures of synthetic ascr#2, ascr#3 and ascr#4 confirmed this hypothesis. A physiological mixture of 20 fmol ascr#2, 20 fmol ascr#3 and 1 pmol ascr#4 reproduced the activity in fractions A and F and elicited significantly stronger male attraction than a mixture of 20 fmol ascr#2 and 20 fmol ascr#3 alone (Fig. 2d). The same mixture of ascr#2, ascr#3 and ascr#4 was then tested on different species of the *Caenorhabditis* genus. *Caenorhabditis brenneri* and *Caenorhabditis remanei* males responded to the mixture in a way very similar to *C. elegans*, whereas *Caenorhabditis briggsae* and *Caenorhabditis japonica* males responded only weakly (Supplementary Fig. 9)[11].

We then tested whether attraction by the identified pheromone components is sex specific. Ascr#3, tested at the concentration that elicits maximal male attraction, was significantly less attractive to hermaphrodites (Supplementary Fig. 10a). Similarly, ascr#2 showed little or no activity for hermaphrodites (Supplementary Fig. 10a). At higher concentrations of ascr#2 and ascr#3, hermaphrodites were strongly deterred whereas males showed neither attraction nor deterrence (Supplementary Fig. 10a). These results demonstrate that the identified ascarosides are sex-specific attractants.

These results indicate that ascarosides regulate both dauer formation and male attraction in *C. elegans*. To determine how these signalling molecules elicit such different biological responses, we compared their activity profiles with regard to both dauer formation and male attraction. We found that ascr#3 elicited the strongest response for male attraction. Males spent approximately ~6.6 times longer in the ascr#3-spotted region than in the control region, whereas ascr#2 elicited a maximum ~2.8-fold increase (Fig. 2b and Supplementary Fig. 10a). However, ascr#4 and the first-published dauer-inducing ascaroside, ascr#1 (ref. 12), the second of which we did not detect in any of our active fractions, were not active in the attraction assay at the range of concentrations tested (data not shown). In the dauer formation assay, ascr#4 was weakly active and ascr#2 and ascr#3 showed strong dauer-inducing activity at the concentrations previously reported (Supplementary Fig. 11a–c). Ascr#3 is much more potent as a male attractant than ascr#2, whereas in the dauer assay ascr#2 is slightly more potent than ascr#3. Together with the observed synergy, these activity profiles suggest that ascr#2 and ascr#3, and possibly ascr#4, act by means of different molecular mechanisms[13,14].

To further investigate the connection between dauer and mating signals, we analysed media extracts of *daf-22* mutants, which are known not to contain the dauer pheromone[15]. We found that neither *daf-22* media extracts nor *daf-22* worm pellet contained ascr#2, ascr#3 or ascr#4, and that *daf-22* media extracts did not attract males at the range of dilutions tested (Supplementary Fig. 12). Additional analyses of extracts from *Escherichia coli* cultures (HB101 and OP50) by LC-MS and NMR spectroscopy confirmed that *E. coli* do not produce ascr#2, ascr#3 or ascr#4 (Supplementary Figs 13–15); however, it is conceivable that bacterial food sources contribute a precursor to the biosynthesis of these compounds.

Given the nature of the attraction assay used in this study (Fig. 1a), the exact concentration of the compounds tested in the scoring region was not well defined. However, it seems likely that the test sample volume (1 μl), when added to the scoring region, was diluted by diffusion, suggesting that the actual concentrations of the assayed compounds on the plate were much lower than the original concentrations in the test sample volumes. Male *C. elegans* change direction of motion more frequently in the presence of an attractant, which correlates with an increase in time spent in the sample scoring region[5], suggesting that reversal frequency could be used as a measure of pheromone perception. Thus, we monitored reversal frequency of males on agar plates with a range of concentrations of the most active pheromone components, ascr#2 and ascr#3, using the automated worm tracking system[16]. As shown in Fig. 3a, reversal frequency is increased by concentrations of ascr#3 as low as 1 pM. For ascr#2, only weak increases of reversal frequency were observed (Supplementary Fig. 10c, d). These results suggest that ascr#3 acts as a male attractant at concentrations more than 10,000 times lower than those required for dauer induction.

General sensory mutants such as *osm-6*, which is expressed in all ciliated neurons in hermaphrodites and males[17], are defective in response to ascr#3 (Fig. 3b). Mutants defective in *osm-3*, which is expressed in a subset of these ciliated sensory neurons and in the male-specific cephalic companion (CEM) neurons[18,19], were also defective in response to ascr#3 (Fig. 3b). Two G-protein α-subunits (*gpa-2* and *gpa-3*) responsible for sensing the dauer pheromone are expressed in the amphid single-ciliated sensory neuron type I (ASI), the amphid double-ciliated sensory neuron type L and the amphid single-ciliated sensory neuron type K (ASK)[20]. Ablation of the ASK neurons, but not the ASI neurons, partially affected response to ascr#3

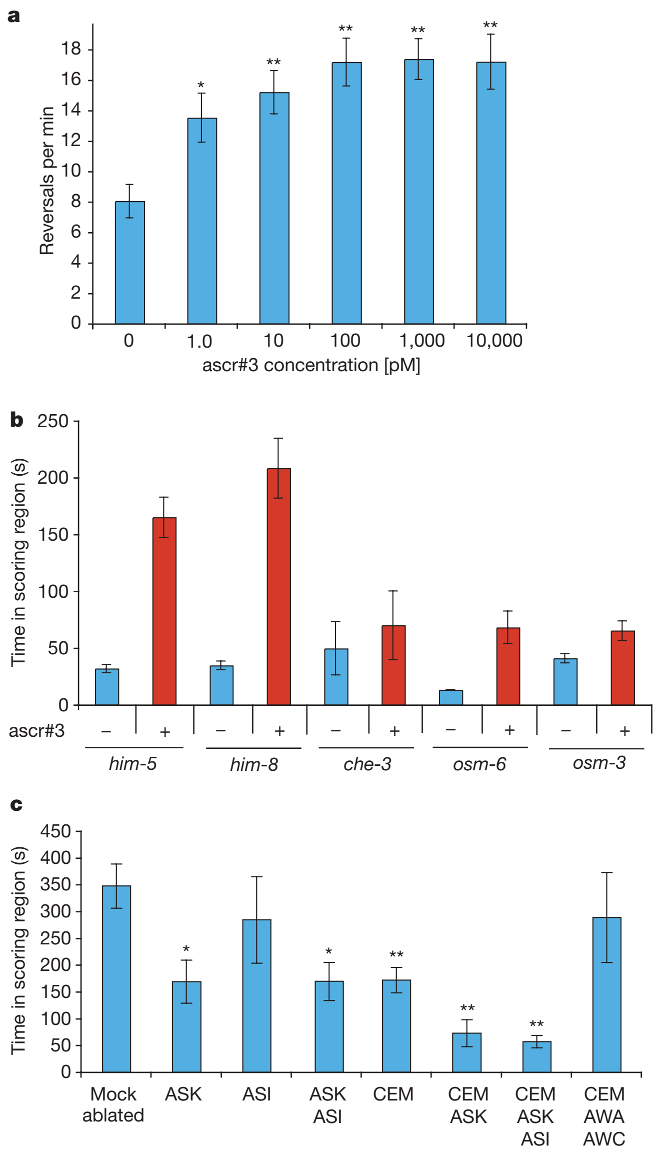


Figure 3 | Neurons mediating response to ascr#3. a, Dose–response curve for male reversal rates on plates with increasing concentrations of ascr#3. $n \geq 15$ animals for each histogram. Error bars, s.e.m.; $*P < 0.05$, $**P < 0.01$, one-factor analysis of variance with post test. **b**, Male attraction by ascr#3 in sensory-deficient mutants[17–19]. Error bars, s.e.m. **c**, General sensory neurons and sex-specific neurons mediate response to ascr#3. Ablation of neurons involved in volatile chemotaxis (amphid winged sensory neuron type A (AWA) and amphid winged sensory neuron type C (AWC)) together with the CEM neurons[6] did not affect response to ascr#3, in comparison with animals lacking only CEM neurons. $n \geq 15$ animals for each ablation set. Error bars, s.e.m.; $*P < 0.05$, $**P < 0.01$, one-factor analysis of variance with post test.

(Fig. 3c). The male-specific CEM neurons have been implicated in sensing the mating signal along with other neurons[6,11]. Removal of the CEM neurons also resulted in partial insensitivity to ascr#3, but removal of the CEM and ASK neurons resulted in complete loss of sensitivity to ascr#3 (Fig. 3c), indicating that response to ascr#3 is mediated by both sex-specific and general sensory neurons. This finding provides a cellular mechanism by which males and hermaphrodites respond differentially to the ascarosides.

Both sexual reproduction and dauer formation, a population-control mechanism that increases larval lifespan and resilience, are major life-history traits. In many organisms, including *C. elegans*, both experimental manipulation and natural genetic variation often have opposite effects on fecundity and lifespan, suggesting a pervasive, inverse relationship between these two traits[21–23]. The discovery that largely overlapping families of small molecules regulate these traits (Supplementary Fig. 1) provides a direct linkage between the corresponding molecular pathways. Characterization of the ascaroside receptors and their downstream targets, as well as the elucidation of ascaroside biosynthesis and the molecular identity of *daf-22*, could provide further insights into how developmental and reproductive pathways are connected.

METHODS SUMMARY

Synchronized *C. elegans* (N2 Bristol) were grown on S-complete medium supplemented with *E. coli* (strain HB101) to desired life stage, washed with M9 buffer to remove bacteria and incubated for 1 h in double-diluted water (ddH_2O) to collect worm-secreted metabolites. Metabolite samples thus produced were tested for mating activity, chromatographically fractionated and analysed using NMR spectroscopy and mass spectrometry (see Methods and Supplementary Information for details).

Mating assays were performed as described previously[5] but were population based. All assays were conducted on plates containing nematode growth medium with a thin film of *E. coli* (OP50) spread throughout the plate as a food source. The worms were given a choice of worm metabolite fraction (or synthetic ascarosides) and control water, and the amount of time spent in each region was measured (see Methods and Supplementary Methods for details). To analyse locomotory behaviour of worms in presence of the ascarosides, standard nematode-growth-medium plates were prepared with the different concentrations of the ascarosides, and worm movement was monitored using an automated tracker to calculate parameters of locomotion[16].

Ascr#1, ascr#2 and ascr#3 were synthesized from L-rhamnose and (2*R*)-propylene oxide (ascr#1, ascr#3) or (2*R*, 5*R*)-hexanediol (ascr#2) as described previously[7], and ascr#4 was subsequently prepared from acetobromo-α-D-glucose and ascr#2 (see Supplementary Methods for details). For NMR spectroscopic comparisons of *daf-22* and wild-type-derived metabolite mixtures, two-week-old liquid cultures of *daf-22* or wild-type (N2) worms raised on *E. coli* (OP50) were extracted and the resulting metabolite samples directly prepared for NMR spectroscopic analyses by means of double-quantum filtered correlation spectroscopy as previously described[7]. For comparison of worm-derived and bacterial metabolites, *E. coli* (OP50) cultures were extracted and subsequently analysed by NMR spectroscopy using the same protocol.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 January; accepted 13 June 2008.
Published online 23 July 2008.

1. Camilli, A. & Bassler, B. L. Bacterial small-molecule signaling pathways. *Science* **311,** 1113–1116 (2006).
2. Dulac, C. & Torello, A. T. Molecular detection of pheromone signals in mammals: from genes to behaviour. *Nature Rev. Neurosci.* **4,** 551–562 (2003).
3. Golden, J. W. & Riddle, D. L. A pheromone influences larval development in the nematode *Caenorhabditis elegans. Science* **218,** 578–580 (1982).
4. Golden, J. W. & Riddle, D. L. A pheromone-induced developmental switch in *Caenorhabditis elegans*: Temperature-sensitive mutants reveal a wild-type temperature-dependent process. *Proc. Natl Acad. Sci. USA* **81,** 819–823 (1984).
5. Simon, J. M. & Sternberg, P. W. Evidence of a mate-finding cue in the hermaphrodite nematode *Caenorhabditis elegans. Proc. Natl Acad. Sci. USA* **99,** 1598–1603 (2002).
6. White, J. Q. *et al.* The sensory circuitry for sexual attraction in *C. elegans* males. *Curr. Biol.* **17,** 1847–1857 (2007).
7. Butcher, R. A., Fujita, M., Schroeder, F. C. & Clardy, J. Small-molecule pheromones that control dauer development in *Caenorhabditis elegans. Nature Chem. Biol.* **3,** 420–422 (2007).
8. Hirsh, D., Oppenheim, D. & Klass, M. Development of the reproductive system of *Caenorhabditis elegans. Dev. Biol.* **49,** 200–219 (1976).
9. Brey, W. W. *et al.* Design, construction, and validation of a 1-mm triple-resonance high-temperature-superconducting probe for NMR. *J. Magn. Reson.* **179,** 290–293 (2006).
10. Carde, R. T. & Elkinton, J. S. *Field Trapping with Attractants: Methods and Interpretation* (eds Hummel, H. E. & Miller, T. A.) 111–129 (Springer, 1984).
11. Chasnov, J. R., So, W. K., Chan, C. M. & Chow, K. L. The species, sex, and stage specificity of a *Caenorhabditis* sex pheromone. *Proc. Natl Acad. Sci. USA* **104,** 6730–6735 (2007).
12. Jeong, P. Y. *et al.* Chemical structure and biological activity of the *Caenorhabditis elegans* dauer-inducing pheromone. *Nature* **433,** 541–545 (2005).
13. Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nature Chem. Biol.* **3,** 408–414 (2007).
14. Lehar, J. *et al.* Chemical combination effects predict connectivity in biological systems. *Mol. Syst. Biol.* **3,** doi:10.1038/msb4100116 (2007).
15. Golden, J. W. & Riddle, D. L. A gene affecting production of the *Caenorhabditis elegans* dauer-inducing pheromone. *Mol. Gen. Genet.* **198,** 534–536 (1985).
16. Cronin, C. J. *et al.* An automated system for measuring parameters of nematode sinusoidal movement. *BMC Genet.* **6,** doi:10.1186/1471-2156-6-5 (2005).
17. Collet, J., Spike, C. A., Lundquist, E. A., Shaw, J. E. & Herman, R. K. Analysis of *osm-6*, a gene that affects sensory cilium structure and sensory neuron function in *Caenorhabditis elegans. Genetics* **148,** 187–200 (1998).
18. Bae, Y. K. *et al.* General and cell-type specific mechanisms target TRPP2/PKD-2 to cilia. *Development* **133,** 3859–3870 (2006).
19. Tabish, M., Siddiqui, Z. K., Nishikawa, K. & Siddiqui, S. S. Exclusive expression of *C. elegans osm-3* kinesin gene in chemosensory neurons open to the external environment. *J. Mol. Biol.* **247,** 377–389 (1995).
20. Zwaal, R. R., Mendel, J. E., Sternberg, P. W. & Plasterk, R. H. Two neuronal G proteins are involved in chemosensation of the *Caenorhabditis elegans* Dauer-inducing pheromone. *Genetics* **145,** 715–727 (1997).
21. Gems, D. & Riddle, D. L. Longevity in *Caenorhabditis elegans* reduced by mating but not gamete production. *Nature* **379,** 723–725 (1996).
22. Partridge, L., Gems, D. & Withers, D. J. Sex and death: what is the connection? *Cell* **120,** 461–472 (2005).
23. Tissenbaum, H. A. & Ruvkun, G. An insulin-like signaling pathway affects both longevity and reproduction in *Caenorhabditis elegans. Genetics* **148,** 703–717 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Human Frontiers Science Program (A.S.E., P.W.S. and P.E.A.T.), a US National Institutes of Health grant (P41 GM079571) to F.C.S., and the Howard Hughes Medical Institute, of which J.S. is an associate and P.W.S. an investigator. NMR data were collected in the UF-AMRIS facility; we thank J. Rocca for assistance. We thank E. Peden and D. Xue for the *ceh-30* strains, C. J. Cronin and A. Choe for advice on behavioural assays, L. R. Baugh for liquid-culture dauer formation assays, B. Fox for assistance with the synthesis of ascr#2, ascr#3 and ascr#4, and M. de Bono, A. Dossey and M. Stadler for discussions. E. Hallem, J. Bungert and D. Hutchinson provided comments on the manuscript.

Author Contributions J.S. and P.W.S. designed the biological experiments and J.S. performed all the biological experiments; F.K. developed the procedure for collecting secreted worm metabolites and designed the chemical experiments and fractionation; F.K. and R.A. produced worm-conditioned water and performed chromatography; F.K., F.C.S., R.U.M., C.Z. and A.S.E. performed structure elucidation by NMR; H.T.A. performed structure elucidation by LC-MS; F.C.S. synthesized ascr#2, ascr#3 and ascr#4; and J.S., F.K., F.C.S., A.S.E., P.E.A.T. and P.W.S. analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.W.S. (pws@caltech.edu), F.C.S. (fs31@cornell.edu) or A.S.E. (art@mbi.ufl.edu).

METHODS

Collecting *C. elegans*-secreted metabolites. Synchronized *C. elegans* (N2 Bristol) with a worm density of 10,000 worms per millilitre was grown at 22 °C at 250 r.p.m. in an incubator shaker in S-complete medium supplemented with *E. coli* (strain HB101): 1% for L2, 2% for L3, 3% for L4, 3% for young adult, 4% for adult and 0% for L1, which was not fed. After worms reached the desired life stages, they were exposed to several wash and filtration (10-μm NITEX nylon filters) steps using M9 buffer to remove bacteria. The worms were collected between the washes either by gentle centrifugation at 121*g* for 30 s or by allowing the worms to settle for 10 min. To remove the bacteria in the gut of the worms, they were placed in M9 buffer in an incubator shaker for 30 min at 22 °C at 250 r.p.m., which was followed by three washes with ddH_2O. Subsequently, *C. elegans*-secreted metabolites were collected by incubating in ddH_2O in an incubator shaker for 1 h at 22 °C at 250 r.p.m. with a worm density of ~30,000 worms per millilitre for L2, L3, young adult and adult; ~15,000 worms per millilitre for L4; and ~100,000 worms per millilitre for L1. The worms were removed from conditioned water by gentle centrifugation at 121*g* for 10 s. The conditioned water was filtered through a 0.2-μm filter, lyophilized and stored at −80 °C. At least three independent experiments were done for each developmental stage. We developed a working unit called 'worm equivalents' to keep track of relative concentrations of unknown compounds. One worm equivalent is the volume of worm water that contains the compounds secreted by one worm in 1 h.

Mating (male-attraction) assay. We modified the single-worm response assay in *C. elegans*[3] to test multiple worms. Standard nematode-growth-medium plates (5-cm diameter) were used for assaying biological activity of the worm-conditioned water. The assay plates consisted of a thin lawn of an *E. coli* OP50 culture, grown overnight, with a ~0.25-cm gap between the bacterial lawn and the edge of the plate to prevent the animals from escaping. Plates were stored at room temperature (20 °C) for two days before being used in trials. Two spots (5-mm diameter) were spotted 1.6 cm apart on a template and stuck to the bottom of the assay plate (Fig. 1a). 0.8 μl of the control and the worm metabolite were placed in the two circles and allowed to dry for approximately 30 s. To remove any bias, control and conditioned water spots were interchanged after every trial. Males and hermaphrodites were harvested daily at the L4 stage and stored at 20 °C overnight with, per plate, 50–60 worms of the same sex to be used as young adults the following day. Five worms were placed ~1.0 cm away from each spot (10 worms total) and allowed to acclimatize for 5 min. Trials were videotaped at 30 frames per second for 15 min using the Unibrain Fire-i software. For each sample, a minimum of four or five trials were conducted each day and each stage was tested on at least three different days.

Purification of mate-finding pheromones. Secreted metabolites were collected from 4,000,000 young adult worms using the method described above. The purification involved a series of fractionation steps guided by the male-attraction assay. Conditioned young adult water was lyophilized and the residue suspended in H_2O. Reverse-phase solid-phase extraction was performed using Sep-Pak Plus C18 cartridges (Waters). The column was eluted sequentially with 50% and 90% MeOH. The active 50%-MeOH fraction was further fractionated by using a SAX anion-exchange column (Alltech) coupled to a SCX cation-exchange column (Alltech). After the neutral fraction was collected, the cation and anion columns were detached and eluted separately with 250, 500 and 1,000 mM KCl. The 250-, 500- and 1,000-mM-KCl fractions were desalted using C18 columns (Waters). The neutral fraction (fraction A) was lyophilized and re-suspended in 6 μl of D_2O containing 0.25 mM of the proton reference standard 3-(trimethylsilyl)propionic acid-D4, for characterization by means of two-dimensional NMR spectroscopy including double-quantum filtered correlation spectroscopy, total correlation spectroscopy, heteronuclear single-quantum coherence, heteronuclear multiple-bond correlation and nuclear Överhauser enhancement spectroscopy. All NMR spectra were acquired at 27 °C using a 1-mm triple-resonance high-temperature superconducting probe and a 600-MHz Bruker Avance II spectrometer[11] (Supplementary Figs 1–6). Total sample amount for these analyses corresponded to about 4,000,000 worm equivalents. Fractions A, F and G were analysed further by LC-MS (Supplementary Fig. 7). In addition, the peak corresponding to ascr#4 in fraction A was analysed by high-resolution mass spectrometry using an Agilent 6210 mass spectrometer: mass of sodium adduct of molecular ion $[M + Na]^+$ calculated for $C_{18}H_{32}O_{10}$ Na, 431.1888 AMU; found, 431.1907 AMU.

LC-MS analysis of ion-exchange fractions. A Thermo Finnigan LCQ Deca XP Max was used with electrospray ionization in positive or negative ion mode in the 50–1,000 AMU range (sheet gas, 25 arbitrary units; sweep gas, 5 arbitrary units; spray voltage, 5.00 kV; capillary temperature, 285 °C; capillary voltage, 3.0 V). Daughter ion spectra were obtained from a dependent scan of the most intense ion in a predefined mass range. The Thermo Separation spectra HPLC system consisted of a P4000 quaternary pump, an AS 3000 autosampler and a UV6000 diode array detector. The tertiary solvents are consisted of methanol with 0.05% formic acid (a), water with 10 mM ammonium formate (b) and 90% acetonitrile–10% water with 10 mM ammonium formate (c). With the column temperature maintained at 60 °C and a solvent flow of 1.0 ml min^{-1}, the C_{18} column (ODS-AMQ, S-5 μm, 20 nm, 250 × 4.6 mm i.d., YMC) was eluted with a solvent composition starting with 4:90:6 (a:b:c) for 2 min followed by a gradient to 4:0:96 in 14 min and then kept at that composition for 5 min. Ultraviolet absorption was monitored at 190 to 400 nm and the solvent flow between the ultraviolet detector and the mass spectroscopy electrospray interface split 9:1 with a low-volume micro needle P450 splitter valve (Upchurch Scientific), making it possible to obtain spectra of eluted compounds and simultaneously collect 90% of the injected material for bioassays.

Vol 454 | 28 August 2008 | doi:10.1038/nature07211

Metabolic gene regulation in a dynamically changing environment

Matthew R. Bennett[1,2]*, Wyming Lee Pang[1]*†, Natalie A. Ostroff[1], Bridget L. Baumgartner[1], Sujata Nayak[1], Lev S. Tsimring[2] & Jeff Hasty[1,2]

Natural selection dictates that cells constantly adapt to dynamically changing environments in a context-dependent manner. Gene-regulatory networks often mediate the cellular response to perturbation[1–3], and an understanding of cellular adaptation will require experimental approaches aimed at subjecting cells to a dynamic environment that mimics their natural habitat[4–9]. Here we monitor the response of *Saccharomyces cerevisiae* metabolic gene regulation to periodic changes in the external carbon source by using a microfluidic platform that allows precise, dynamic control over environmental conditions. We show that the metabolic system acts as a low-pass filter that reliably responds to a slowly changing environment, while effectively ignoring fast fluctuations. The sensitive low-frequency response was significantly faster than in predictions arising from our computational modelling, and this discrepancy was resolved by the discovery that two key galactose transcripts possess half-lives that depend on the carbon source. Finally, to explore how induction characteristics affect frequency response, we compare two *S. cerevisiae* strains and show that they have the same frequency response despite having markedly different induction properties. This suggests that although certain characteristics of the complex networks may differ when probed in a static environment, the system has been optimized for a robust response to a dynamically changing environment.

To probe the response of a metabolic gene network to a fluctuating environment, we developed a microfluidic platform that can subject a population of cells to a continuously varying supply of medium (Fig. 1). The device is designed to generate a fluctuating signal of medium by dynamically combining two medium-filled reservoirs in accordance with a time-dependent function. Feeding channels deliver the media downstream to a customizable growth chamber, which for this study was constructed to constrain a population of yeast cells to grow in a monolayer, permitting long-term data acquisition[10]. The composition of the medium is dynamically controlled by a fluidic switch[11], such that changes in the upstream source may be detected almost immediately by the cells. The fluidic switch was optimized to generate a linear range of mixing ratios from the two media inputs, allowing a variety of periodic waveforms or random signals to be generated (see Supplementary Information for full details of the device).

As a quantifiable reporter of the cellular response to environmental fluctuations, we fused the native Gal1 protein of *S. cerevisiae* to the yeast-optimized enhanced cyan fluorescent protein (yECFP)[12,13]. The enzymes for galactose utilization, including Gal1, are among the most tightly regulated proteins in yeast. Because glucose requires much less energy to metabolize, cells will consume galactose only if glucose is not available. *S. cerevisiae* has therefore evolved a highly complex regulatory network to ensure that the galactose-metabolizing enzymes will be strongly activated when they are needed, but tightly repressed if glucose is present in the environment (Fig. 2a). Because the network has been well studied and involves regulatory motifs common to many higher organisms, galactose utilization is a model for gene regulation. To build on the current understanding of its robust regulatory mechanisms, we employed our microfluidic platform to monitor the dynamics of network activation and repression in response to sinusoidal perturbations of glucose over a galactose background.

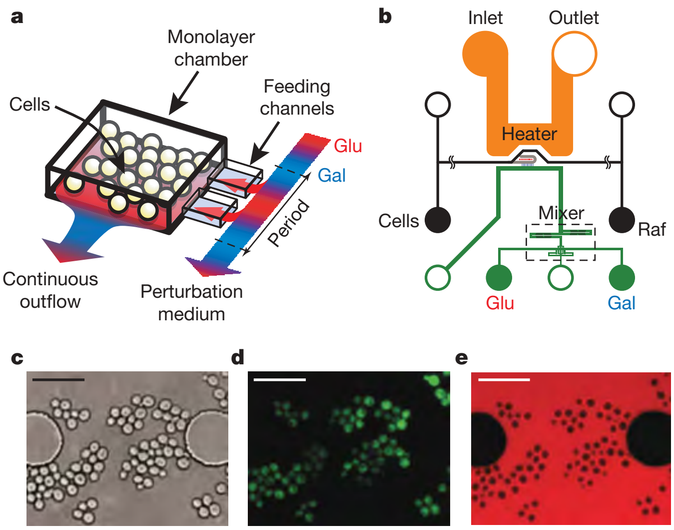


Figure 1 | Design and implementation of the microfluidic platform developed for our study. **a**, Conceptual design of the imaging chamber. The chamber is coupled to the switch output channel by means of multiple 'feeding' channels 1 μm tall. The feeding channels are fed by a controllable waveform generator that creates sinusoidal perturbations in the glucose concentration while maintaining constant background levels of galactose. **b**, An overview of the design showing the layout of the device. The device makes use of three flow networks for loading cells (middle, black), generating microenvironmental waveforms (bottom, green), and controlling on-chip temperature (top, orange). The imaging chamber (centre, grey region) is designed to be about 4 μm tall to constrain a population of yeast cells to grow in a monolayer. **c**, Representative bright-field image of cells growing in the imaging chamber. These images were used to measure the total size of the colony. Large circles are support posts in the chamber. Scale bar, 25 μm. **d**, Green fluorescence image of the same cells as in **c**. These images allowed us to measure the amount of Gal1 in each cell. **e**, Red fluorescence image of the chamber. The glucose medium also contained a red fluorescent dye; the intensity of the red fluorescence was therefore proportional to the amount of glucose in the chamber at any given time.

[1]Department of Bioengineering, and [2]Institute for Nonlinear Science, University of California, San Diego, La Jolla, California 92093, USA. †Present address: Institute for Systems Biology, Seattle, Washington 98103, USA.
* These authors contributed equally to this work.

A population of yeast cells was subjected to sinusoidal waves of glucose concentration over a 0.2% (w/v) galactose background, with glucose concentration varying from 0% (no repression of *GAL1* transcription) to 0.25% (full repression; see Supplementary Information for repression data). For each run we changed the frequency of the glucose signal, varying the period from 0.75 to 4.5 h, and we imaged the population for a minimum of four full cycles. Time-lapse fluorescence imaging of the cell population in the growth chamber was used to calculate the amplitude ratio and phase shift of the cellular response relative to the glucose signal. The results show a maximum response frequency of about 5.6 radians h^{-1} (1.125 h period). At this frequency, the response trace was indistinguishable from a normal step-function response, whereas at the lower frequencies the temporal fluorescence trajectories clearly oscillated in response to the signal. In this sense, the galactose system seems to function as a low-pass filter that reliably responds to a slowly changing environment, while effectively ignoring fluctuations that are too fast for the cell to mount an efficient response.

Because the sinusoidal driving of the galactose utilization network leads to complex cellular behaviour, we used computational modelling to simulate the response and to uncover key aspects of the network architecture that give rise to the observed behaviour[14]. In particular, we were interested in how the interplay of the galactose and glucose utilization networks gives rise to the observed frequency response to carbon source fluctuations. By itself, the turnover of Gal1–yECFP, due either to dilution or to active degradation (or both), leads to low-pass filtering of periodic signals. However, feedback loops inherent in gene regulatory networks can alter the response of proteins to stimuli[15]. Therefore, to simulate the effects of galactose activation and glucose repression on our experimental data, we adapted a comprehensive model of the galactose network described previously[16]. This model includes the transcription and translation of the *GAL1*, *GAL2*, *GAL3*, *GAL4* and *GAL80* genes as well as the interactions of their respective proteins with each other and with galactose (such as dimerization, transport and metabolism). Whenever possible we used parameter values either at or close to the values reported previously[16]. In addition to this galactose network model, it was necessary to model the dynamics of the glucose network. The glucose network is much more complex than that of galactose[17–19] and models for it are much less well established. We therefore chose to model the glucose network with a simplified module describing a basic transport regulatory system. In it, protein products of the glucose network are responsible for transporting external glucose into the cell while internalized glucose acts to induce transcription in the network, giving rise to a positive feedback loop (see Fig. 2a).

Calibration of the computational model to the experimental data led to several important observations that would not have arisen from an analysis of steady-state batch culture data. The large amplitude ratios observed at low frequencies suggested that when glucose was added to the system the degradation rates of some galactose network components were greater than in the absence of glucose. Previous studies have suggested that components of the glucose network can actively degrade messenger RNA produced by genes involved in the galactose/glucose switch[8], and this phenomenon has also been shown to exist for the mRNA of other genes[20–22]. We therefore added enzymatic decay terms (governed by Michaelis–Menten dynamics) to the equations describing the dynamics of the *GAL1* and *GAL3* mRNA and found that it greatly increased the accuracy of the model. These two genes are among those in the galactose network that are targeted by the glucose-induced protein Mig1, which represses transcription by binding to upstream regulatory sites[19]. Thus, if proteins from the glucose network do actively degrade galactose network transcripts, *GAL1* and *GAL3* are likely targets. To test this prediction, we measured the degradation rates of *GAL1* and *GAL3* in both galactose and glucose. Both transcripts showed a 2–8-fold increase in their decay rates in the presence of glucose (see Fig. 2b and Supplementary Information), which is consistent with the values predicted by the computational model. This form of post-transcriptional regulation, in which glucose acts to downregulate Gal protein synthesis, is a previously unknown source of regulation in the galactose utilization network. Furthermore, the inclusion of glucose-mediated mRNA decay results in a model that accurately reproduces the dynamic response of a population of cells to sinusoidal repression over a large range of frequencies (Fig. 3).

Batch-culture induction characteristics for metabolic genes can vary from strain to strain or depend sensitively on the growth state of the culture. We were therefore also interested in using the model to determine how galactose induction differences would affect the response to the glucose fluctuations. The model demonstrated that

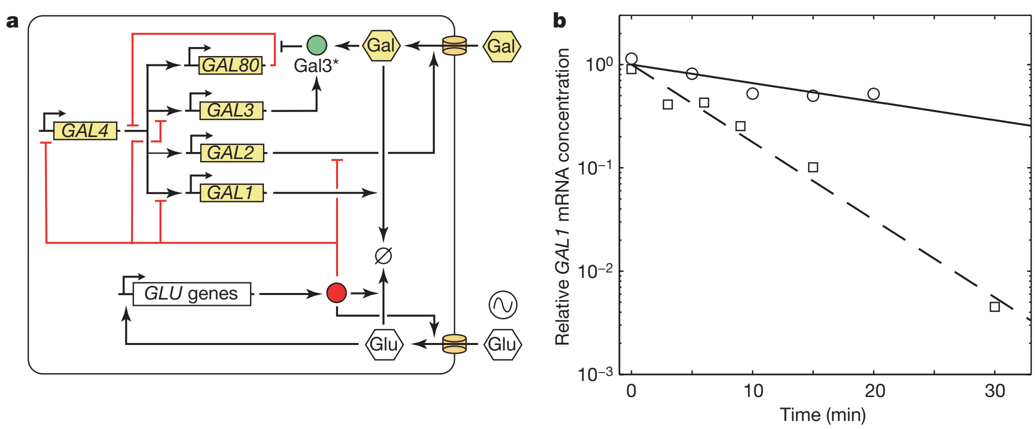


Figure 2 | Regulation in the galactose utilization network. **a**, Diagram of the gene regulatory networks involved. The regulatory genes in the galactose network are activated by the Gal4 protein, which binds to upstream activation sites. The *GAL80* gene provides negative feedback in the system by prohibiting the inducing affects of Gal4. Positive feedback is provided by both *GAL2* and *GAL3*. Internalized galactose can bind to Gal3, and the resulting complex binds to Gal80. Gal80 bound to the Gal3–galactose complex is incapable of repressing Gal4. In addition, the transporter Gal2 increases the amount of internal galactose, which stimulates the galactose network. The glucose network inhibits the transport of galactose and represses transcription of the galactose network in the presence of glucose through the action of Mig1, which can bind to upstream regulatory sites of *GAL1*, *GAL3* and *GAL4* (ref. 19). The glucose network also regulates the hexose transporter genes (*HXT*), responsible for transporting glucose into the cell[27], which then activates the glucose network. **b**, Experimentally measured decay of *GAL1* transcripts in galactose (circles) and glucose (squares). Also shown are the best-fit lines corresponding to half-lives of about 17 min in galactose (solid line) and 4 min in glucose (dashed line), which are similar to the values predicted by the numerical model. Data are normalized to the initial concentration of mRNA predicted by the best-fit lines. Similar results for *GAL3* transcripts are shown in Supplementary Information.

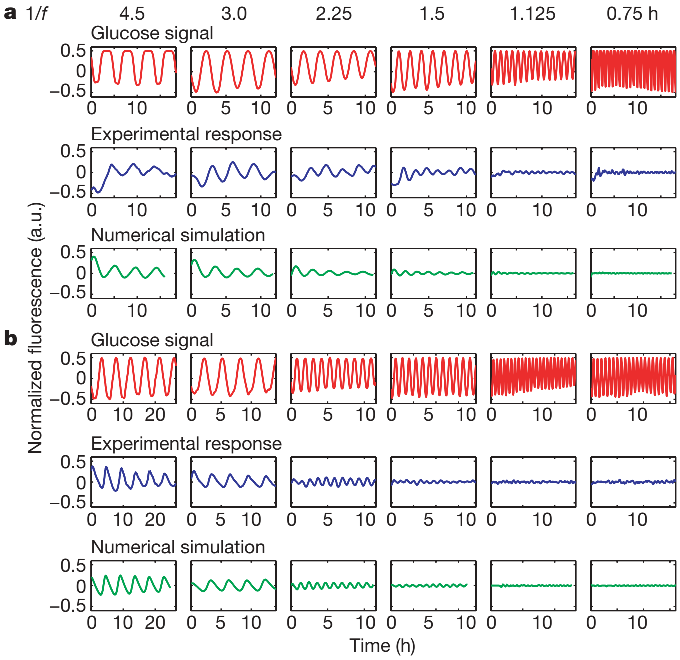


Figure 3 | Experimental and computational results for cells of two yeast strains expressing a *GAL1–yECFP* fusion gene in response to alternating glucose and galactose media. **a**, Strain K699; **b**, strain YPH499. The top row for each strain depicts the input glucose signal that was measured during each experimental run and was also used to simulate the responses. The mean fluorescence of a red tracer dye, representing the glucose concentration in the medium, is normalized and subtracted from 1 to represent the 'induction' signal used in the experimental and computational runs above. The middle rows show normalized and detrended fluorescence trajectories for a population of cells as they respond to glucose waves of various frequencies over a galactose background. In the absence of glucose, galactose induces the transcription of *GAL1–yECFP*, causing an increase in cellular fluorescence. However, as glucose is introduced into the extracellular environment, transcription of the galactose enzymes is shut off, causing a decrease in fluorescence signal as the Gal1–yECFP protein is degraded. Oscillation periods are (from left to right) 4.5, 3.0, 2.25, 1.5, 1.125 and 0.75 h. For input waves with a period shorter than 1.125 h, cells no longer responded to sinusoidal repression in a periodic fashion, demonstrating their ability to 'filter' out high-frequency environmental fluctuations. The bottom rows show simulation results for the same frequencies as above. The model, calibrated to experimental induction and repression data, accurately reproduces the cellular responses over a large range of frequencies.

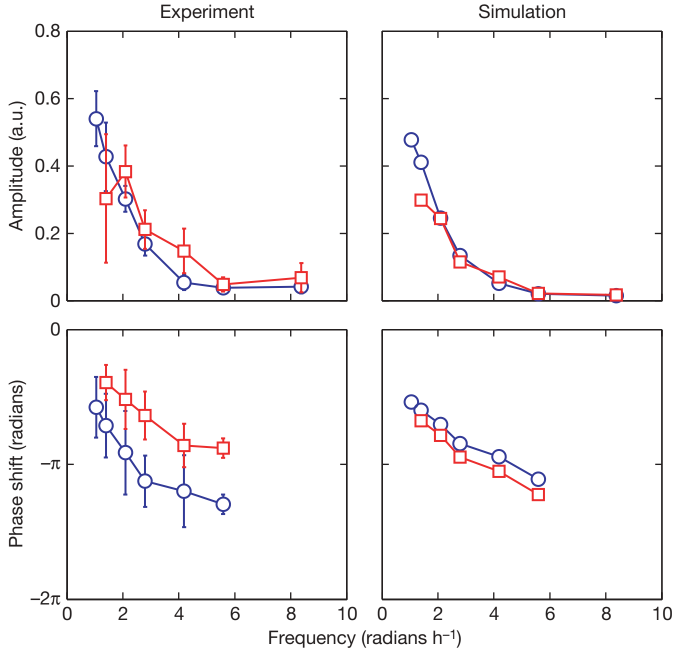


Figure 4 | Experimental and computational comparison of two yeast strains. The amplitude (top row) and phase shift (bottom row) of the response of cells to sinusoidal repression at various frequencies are shown for both K699 (red) and YPH499 (blue) strains (error bars represent s.d.). Strain YPH499 is known to have a deficiency in the galactose utilization network. For the highest-frequency trial, reliable phases could not be calculated because of noise; these results have been omitted from the graphs. The experimental data (left column) show that the amplitude responses of the two strains are strikingly similar, especially considering their significantly different induction curves (see Supplementary Information). This phenomenon was predicted by model simulations, because slight modifications to the model parameters that affected the induction and repression curves did not affect the cell population's robust response to a dynamic environment. This suggests that the complex structure of the glucose and galactose networks may confer robustness on cells even if faced with seemingly detrimental network deficiencies. The phase responses (bottom row) of the two strains showed a marked difference, with YPH499 cells having a greater phase lag than K699 cells.

significantly different galactose induction does not necessarily lead to significant differences in the response characteristics (data not shown). In other words, the model led to the hypothesis that deficiencies in network induction capabilities might not hinder a cell's ability to adapt and thrive in a changing environment. The yeast strain used to collect the present data, K699, is sensitive to external galactose concentrations, with full induction of the galactose network occurring at about 0.05% (w/v) galactose. To test our hypothesis, we turned to a strain (YPH499) that is known to have a deficiency in the galactose utilization network that causes it to require more galactose than 'normal' to induce production of the galactose enzymes[23]. YPH499 is a derivative of a *GAL2* mutant strain, and although the mutations were reportedly repaired, the *GAL2* alleles in many of the derivative strains have been shown to cause significantly impaired galactose uptake[23]. Gal2 protein is responsible for the transport of extracellular galactose into the cell and its activity is markedly different in YPH499 from that in K699. Our flow cytometry population data demonstrated that YPH499 cells require about tenfold more galactose than K699 cells to reach full induction (see Supplementary Information).

Despite the difference in induction sensitivity between K699 and YPH499 cells, our model predicted that inefficient Gal2 transport does not translate into a less robust response to a fluctuating environment. This suggests that the complex interplay of the glucose and galactose networks may confer robustness on cells even if faced with deficiencies in the induction characteristics. To validate this finding, we repeated the microfluidic runs at each frequency, this time using the YPH499 strain with a Gal1–yECFP fusion. As predicted, the amplitude responses of the two strains are strikingly similar (Fig. 4), especially considering the significant difference in their galactose sensitivity. We do not at present know the underlying mechanistic property of the regulatory network that leads to the robust response of the two strains. Future studies might endeavour to deduce this mechanism through the systematic deconstruction of the regulatory elements in a single strain. Although the present study shows how robustness can occur despite large differences in induction characteristics, one could further investigate the generality of this phenomenon by comparing the responses of many different strains to different types of temporal perturbation.

METHODS SUMMARY

Dynamic environment experiments. Cells containing a *GAL1–yECFP* fusion were imaged every 5 min for up to 24 h by using time-lapse fluorescent microscopy to estimate the concentration of Gal1 as a function of time. Cells were constrained to grow in a custom-designed microfluidic platform that permitted the dynamically controlled mixing of two growth media. Our inducing medium contained

2% raffinose and 0.2% galactose, whereas the repressing medium contained 2% raffinose, 0.2% galactose and 0.25% glucose. The resulting images were processed with cell segmentation and tracking software, and the population-averaged fluorescence concentrations were measured. To ensure correct waveform generation, glucose concentrations were monitored by introducing a red fluorescent tracer dye (sulphorhodamine 101, 0.01 mg ml^{-1}) to the repressing medium.

Microfluidic chips and waveform generation. The polydimethylsiloxane microfluidic devices were designed to permit the monolayer growth of yeast cells in the imaging chamber and were fabricated with standard replica-moulding techniques[24–26]. An upstream fluidic switch controlled the input of medium into the chamber by mixing the flows of the inducing and repressing media. The mixing ratio of the two media was governed by a software-controlled, custom-designed pressurization system that was able to produce time-varying waveforms consistently.

mRNA degradation experiments. The degradation rates of *GAL1* and *GAL3* transcripts were measured with standard reverse-transcriptase-mediated quantitative polymerase chain reaction techniques. Knockout strains for both genes were first created, and then ectopic *GAL1* and *GAL3* were placed back into the cell under the control of a doxycycline-repressible promoter. Half-lives of mRNAs were measured from cells grown in the presence or absence of glucose.

Complete details of all materials and methods used and the specifics of the computational model are available in Supplementary Information.

Received 14 April; accepted 25 June 2008.
Published online 30 July 2008.

1. Beadle, G. W. & Tatum, E. L. Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl Acad. Sci. USA* **27**, 499–506 (1941).
2. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
3. Douglas, H. C. & Hawthorne, D. C. Regulation of genes controlling synthesis of the galactose pathway enzymes in yeast. *Genetics* **54**, 911–916 (1966).
4. Thattai, M. & Shraiman, B. I. Metabolic switching in the sugar phosphotransferase system of *Escherichia coli*. *Biophys. J.* **85**, 744–754 (2003).
5. Lipan, O. & Wong, W. H. The use of oscillatory signals in the study of genetic networks. *Proc. Natl Acad. Sci. USA* **102**, 7063–7068 (2005).
6. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078 (2005).
7. Kruse, K. & Julicher, F. Oscillations in cell biology. *Curr. Opin. Cell Biol.* **17**, 20–26 (2005).
8. Ronen, M. & Botstein, D. Transcriptional response of steady-state yeast cultures to transient perturbations in carbon source. *Proc. Natl Acad. Sci. USA* **103**, 389–394 (2006).
9. Thattai, M. & van Oudenaarden, A. Stochastic gene expression in fluctuating environments. *Genetics* **167**, 523–530 (2004).
10. Cookson, S., Ostroff, N., Pang, W. L., Volfson, D. & Hasty, J. Monitoring dynamics of single-cell gene expression over multiple cell cycles. *Mol. Syst. Biol.* **1**, doi:10.1038/msb4100032 (2005).
11. Groisman, A. *et al.* A microfluidic chemostat for experiments with bacterial and yeast cells. *Nature Methods* **2**, 685–689 (2005).
12. Sheff, M. A. & Thorn, K. S. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast* **21**, 661–670 (2004).
13. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
14. Hasty, J., McMillen, D., Isaacs, F. & Collins, J. J. Computational studies of gene regulatory networks: *in numero* molecular biology. *Nature Rev. Genet.* **2**, 268–279 (2001).
15. Savageau, M. A. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature* **252**, 546–549 (1974).
16. de Atauri, P., Orrell, D., Ramsey, S. & Bolouri, H. Evolution of 'design' principles in biochemical networks. *Syst. Biol.* **1**, 28–40 (2004).
17. Demir, O. & Kurnaz, I. A. An integrated model of glucose and galactose metabolism regulated by the *GAL* genetic switch. *Comput. Biol. Chem.* **30**, 179–192 (2006).
18. Kaniak, A., Xue, Z., Macool, D., Kim, J. & Johnston, M. Regulatory network connecting two glucose signal transduction pathways in *Saccharomyces cerevisiae*. *Eukaryot. Cell* **3**, 221–231 (2004).
19. Verma, M., Bhat, O. J. & Venkatesh, K. V. Steady-state analysis of glucose repression reveals hierarchical expression of proteins under Mig1p control in *Saccharomyces cerevisiae*. *Biochem. J.* **388**, 843–849 (2005).
20. Scheffler, I. E., de la Cruz, B. J. & Preito, S. Control of mRNA turnover as a mechanism of glucose repression in *Saccharomyces cerevisiae*. *Biochem. Cell Biol.* **30**, 1175–1193 (1998).
21. de la Cruz, B. J., Prieto, S. & Scheffler, I. E. The role of the 5′ untranslated region (UTR) in glucose-dependent mRNA decay. *Yeast* **19**, 887–902 (2002).
22. Andrade, R. P., Kötter, P., Entian, K. D. & Casal, M. Multiple transcripts regulate glucose-triggered mRNA decay of the lactate transporter *JEN1* from *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* **332**, 254–262 (2005).
23. Rohde, J. R., Trinh, J. & Sadowski, I. Multiple signals regulate *GAL* transcription in yeast. *Mol. Cell. Biol.* **20**, 38803886 (2000).
24. Whitesides, G. M., Ostuni, E., Takayama, S., Jiang, X. Y. & Ingber, D. E. Soft lithography in biology and biochemistry. *Annu. Rev. Biomed. Eng.* **3**, 335–373 (2001).
25. Whitesides, G. M. *et al.* Soft lithography and bioanalysis. *Abstr. Pap. Am. Chem. Soc.* **227**, U113–U113 (2004).
26. Xia, Y. & Whitesides, G. M. Soft lithography. *Angew. Chem. Int. Edn Engl.* **37**, 550–575 (1998).
27. Boles, E. & Hollenberg, C. P. The molecular genetics of hexose transport in yeasts. *FEMS Microbiol. Rev.* **21**, 85–111 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Groisman for useful discussions regarding microfluidic design; D. Volfson and C. Grilly for aid in development and testing of image segmentation and tracking algorithms, and M. Ferry for his suggestions on microbiology. This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.H. (hasty@ucsd.edu).

LETTERS

Crystal structure of the polymerase PA_C–$PB1_N$ complex from an avian influenza H5N1 virus

Xiaojing He[1], Jie Zhou[1], Mark Bartlam[2], Rongguang Zhang[3], Jianyuan Ma[1], Zhiyong Lou[4], Xuemei Li[1,4], Jingjing Li[1], Andrzej Joachimiak[3], Zonghao Zeng[1], Ruowen Ge[5], Zihe Rao[1,2,4] & Yingfang Liu[1]

The recent emergence of highly pathogenic avian influenza A virus strains with subtype H5N1 pose a global threat to human health[1]. Elucidation of the underlying mechanisms of viral replication is critical for development of anti-influenza virus drugs[2]. The influenza RNA-dependent RNA polymerase (RdRp) heterotrimer has crucial roles in viral RNA replication and transcription. It contains three proteins: PA, PB1 and PB2. PB1 harbours polymerase and endonuclease activities and PB2 is responsible for cap binding[3,4]; PA is implicated in RNA replication[5–10] and proteolytic activity[11–14], although its function is less clearly defined. Here we report the 2.9 ångström structure of avian H5N1 influenza A virus PA (PA_C, residues 257–716) in complex with the PA-binding region of PB1 ($PB1_N$, residues 1–25). PA_C has a fold resembling a dragon's head with $PB1_N$ clamped into its open 'jaws'. $PB1_N$ is a known inhibitor that blocks assembly of the polymerase heterotrimer and abolishes viral replication. Our structure provides details for the binding of $PB1_N$ to PA_C at the atomic level, demonstrating a potential target for novel anti-influenza therapeutics. We also discuss a potential nucleotide binding site and the roles of some known residues involved in polymerase activity. Furthermore, to explore the role of PA in viral replication and transcription, we propose a model for the influenza RdRp heterotrimer by comparing PA_C with the λ3 reovirus polymerase structure, and docking the PA_C structure into an available low resolution electron microscopy map.

PA is an important protein in the polymerase heterotrimer and may be required for replication and transcription of viral RNA (vRNA) and endonuclease cleavage of the cap RNA primer[5–9,15]. It reportedly induces proteolysis of viral and host proteins[11–14] and may also be involved in virus assembly[16]. Recombinant PA can be cleaved into two main fragments by trypsin digestion: a ~25-kilodalton (kDa) fragment containing the amino-terminal region of PA, and a ~55-kDa fragment containing the remainder of the protein[15]. The N-terminal fragment is understood to be sufficient for proteolytic activity and important for RNA synthesis activity of the polymerase complex[11,14,15]. The carboxy-terminal region of PA is thought to bind to PB1 for complex formation and nuclear transport[17–19]. Previous reports have shown that the PB1 N-terminal 25 residues ($PB1_N$) specifically bind to the C-terminal region of PA[2,19,20].

The structure of an N-terminally truncated PA covering residues 257–716 (termed PA_C) was determined in a complex with a known interaction peptide from PB1 ($PB1_N$, residues 1–25, Supplementary Fig. 2B) by X-ray crystallography (see Supplementary Table 1 for statistics). PA_C consists of 13 α-helices, one short 3_{10} helix (η1), nine β-strands and several loops/turns (Fig. 1a, b). PA_C resembles the head of a dragon and can be subdivided into two parts: domain I, the 'brain', and domain II, the 'mouth' (Fig. 1a). The strands β1 to β7 of the brain form a twisted plane surrounded by five α-helices (α1, α2, α3, α6 and α7) and the η1 helix. Anti-parallel strands β6 and β7 are linked by a four-residue turn and extend from the left side of the PA_C brain. The mouth consists of strands β8 and β9, and helices α4, α5 and α8–α12; helices α11 and α13 form the lower jaws, whereas helices α8 and α10 form the upper jaws. Structural comparison using Dali (http://www.ebi.ac.uk/dali) indicated that PA_C has a new fold.

A deep semi-circular groove (Figs 1a and 2a, b) extending down to the core of the brain is ~25 Å in diameter and highly basic, which is favourable for RNA binding. The extended α3–α4 loop consists largely of negatively charged residues and is situated below this deep groove, where it completes the circular structure and forms a 'neck' for the dragon head (Fig. 1a). Residues lining the putative RNA binding groove, including K328, K539, R566 and K574, are highly conserved in all three influenza virus types (Supplementary Fig. 1A, blue stars). The K539A mutation significantly disrupts complementary RNA (cRNA) and vRNA synthesis, suggesting that the groove is important for RNA binding[5].

An elliptical ~14 Å diameter channel, formed by the concave side of the twisted β-sheet and the inside ends of several helices from domain I, lies between the two domains and connects to the putative RNA binding groove (Fig. 2a, b). It is loosely occupied by the N terminus of the PA_C polypeptide, which is too far from either side to form strong interactions with residues lining the channel and may be a crystallographic artefact (Fig. 2b, c). Conserved residues K539, R566, K574 and N696 lie on the surface connecting the groove to the entrance of the channel. Residues E410, K461, E524 and K536 lie on the inside surface of the channel (Supplementary Fig. 1A, blue stars; and Fig. 2a). Mutations of several of these residues are known to influence polymerase activity[5]: E524A impairs RNA synthesis activity and inhibits virus production; E410A decreases polymerase activity; and the K536A/W537A double mutation impairs RNA synthesis activity of the heterotrimer (Fig. 2c)[16]. The precise role of this channel remains to be established, although the highly conserved nature points towards its functional importance.

The R638A mutation promotes generation of interfering RNA, thus giving rise to transcription defects that can be overcome by the C453R mutation (Fig. 2c)[9]. The long side chain of R638 (α10) forms a hydrogen bond and an ion bridge with the carbonyl oxygen and side chain of residue E449 (α5), respectively (Supplementary Fig. 3B). C453 (α6) is about 5 Å from R638 and E449. The R638A mutation would break the interaction between R638 and E449 and cause α5 to move into the proposed RNA binding surface between PA and PB1, thus interfering with RNA synthesis activity of the heterotrimer. The C453R mutation would enable a new interaction with E449 to be formed that would restrain α5, thus at least partially recovering PA

[1]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. [2]College of Life Sciences and Tianjin State Laboratory of Protein Sciences, Nankai University, Tianjin 300071, China. [3]Midwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. [4]Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China. [5]Department of Biological Sciences, National University of Singapore, 117543 Singapore.

from the defect caused by the R638A mutation. The long side chain of C453R may interfere with the normal function of R638, which would explain why a virus containing a single C453R mutation could not be produced. Other mutations known to affect polymerase activity are listed in Supplementary Table 2.

$PB1_N$ binds obliquely between the jaws of PA_C with its N terminus pointing towards the back of the mouth and its C terminus extending outwards (Figs 1a and 2d). Residues from 5 to 11 of $PB1_N$ fold into a short helix (η2). The interaction with $PB1_N$ is largely mediated by α8, α10, α11 and α13 of PA_C; the four helices form a hydrophobic core at the tip of the mouth and tightly interact with $PB1_N$ by means of hydrophobic interactions, hydrogen bonds and van der Waals forces (Fig. 2d). Other minor interactions are provided by α4 and the α9–α10 loop. A short LLFL motif from residues 7–10 of $PB1_N$, known to be important for the interaction with PA[19], interacts with the PA_C hydrophobic core formed by F411 (α4), M595 (α8), L666 (α11), W706 and F710 (α13), and V636 and L640 (α10; Fig. 1a). W706 also interacts with residues V3 and N4; Q408 and N412 (α4) interact with V3 and D2; and Q670 (α11) interacts with $PB1_N$ residues F9, V12, P13 and A14. Residues 620 and 621 on β8 are also located in the $PB1_N$ interaction surface (Supplementary Fig. 1A, green arrows). The W706A/Q670A double mutation disrupts the binding of $PB1_N$ to PA_C (Fig. 2e), as do the L666G/F710E, L666G/F710G and W706A/F710Q double mutations (Supplementary Fig. 4).

$PB1_N$ inhibits influenza A viral replication by interfering with polymerase activity, presumably by blocking assembly of the polymerase heterotrimer[2]. Our structure identifies a very important PB1 binding region of PA, and can thus be used as a basis to design new anti-influenza compounds. Protein–protein interfaces often involve a large surface area, which can present problems for drug discovery. In this case, however, relatively few residues drive the binding of $PB1_N$ to PA_C, indicating that designing small molecule inhibitors

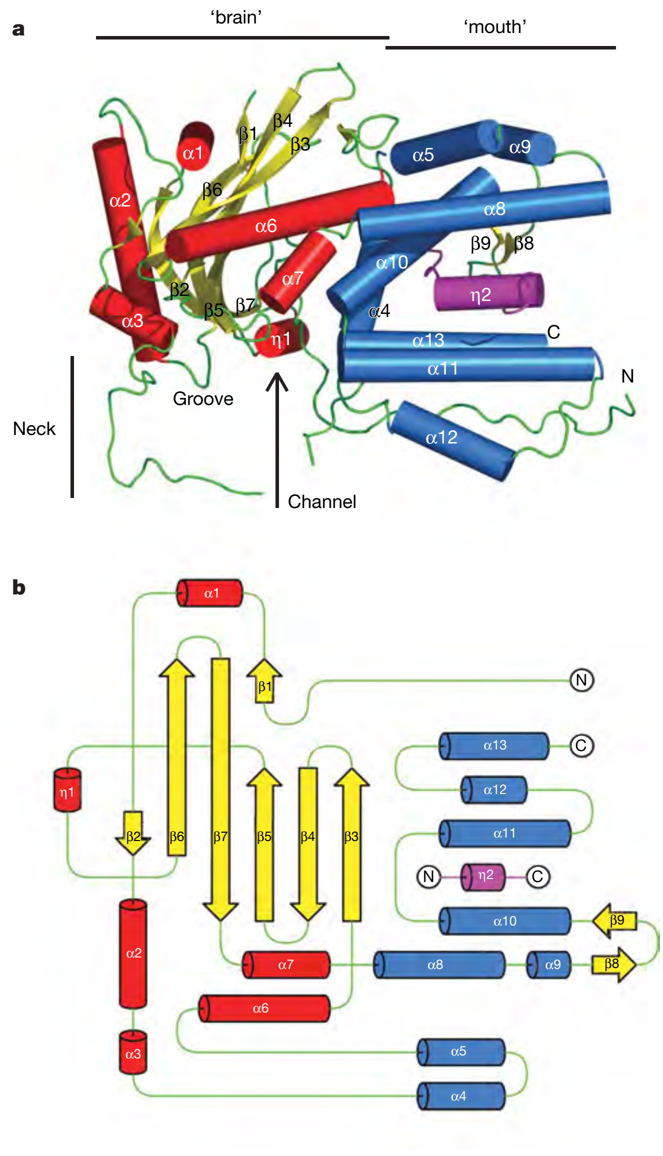


Figure 1 | The PA_C–$PB1_N$ complex structure. **a**, The overall structure of the PA_C–$PB1_N$ complex. The structure is coloured according to secondary structure and elements are labelled. Helices are shown as cylinders and are red in the brain domain and blue in the mouth domain; strands are yellow and loops are green. The $PB1_N$ peptide is magenta. **b**, Topology diagram for the PA_C–$PB1_N$ complex structure. The colouring is consistent with **a**.

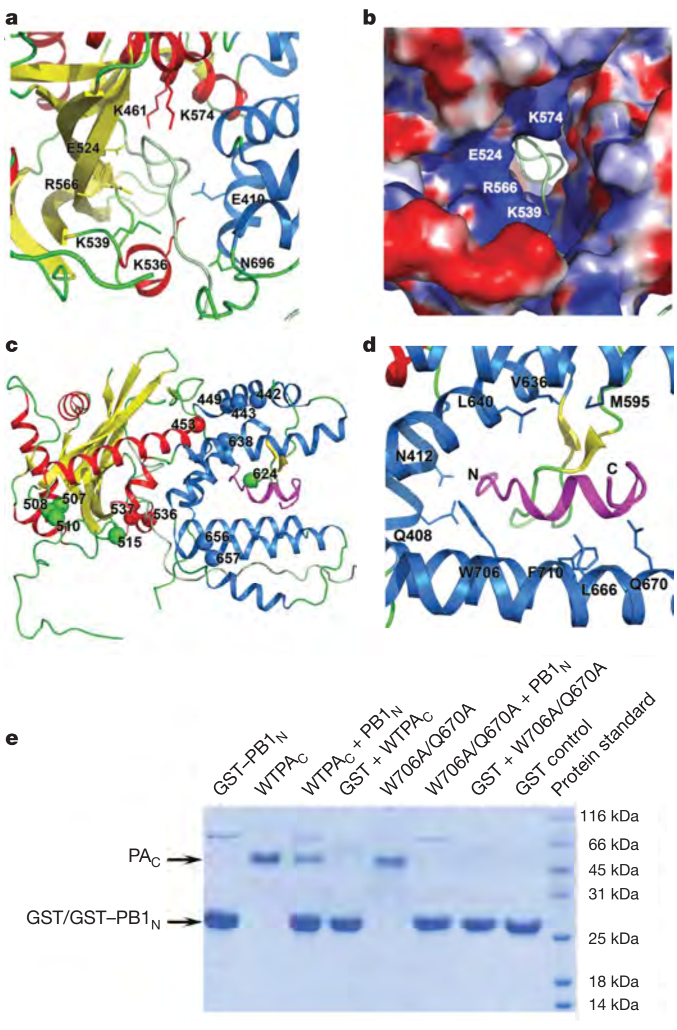


Figure 2 | Representations of the PA_C groove and channel. **a**, An enlarged view of the PA_C groove and channel, rotated by 90° about the *x* axis relative to Fig. 1a and with the same colouring scheme. The extended N-terminal loop from residues 257–288 is coloured pale green. Conserved residues lining the groove and channel are shown in stick representation and labelled. **b**, A surface representation of the same region of PA_C coloured by electrostatic charge from red ($-10\ k_bT/e_c$) to blue ($+10\ k_bT/e_c$), where k_b is the Boltzmann constant, T is temperature and e_c is the electron charge. Residues lining the highly basic groove are labelled. **c**, Location of known mutations in the PA_C structure. PA_C and $PB1_N$ are shown in ribbon representation and coloured according to the scheme in Fig. 1. **d**, Interaction between PA_C and $PB1_N$. PA_C and $PB1_N$ are shown in ribbon representation and coloured according to the scheme in Fig. 1. Residues interacting with the $PB1_N$ peptide are shown in stick representation and labelled. **e**, *In vitro* binding assay for $PB1_N$ to the W706A/Q670A PA_C double mutant. $WTPA_C$, wild-type PA_C.

of this interaction is feasible. Mutation of specific residues in $PB1_N$, including V3, N4, P5, L7, L8, F9 and L10, causes the loss of more than two-thirds of the binding affinity, with significant reductions in polymerase activity and virus production[19]. D2V and A14D mutations in $PB1_N$ do not significantly influence the binding affinity to PA, but they abolish polymerase activity and virus production, suggesting that D2 and A14 have other crucial functions[19]. An L13P mutation in PB1 from a H7N7 virus is associated with increased virulence[21]; in our $PB1_N$ structure from a H5N1 virus, P13 breaks the η2 helix and the associated structural changes may result in increased RNA synthesis by PB1. $PB1_N$ residues interacting with PA are conserved across type A, B and C influenza viruses, and PA residues shown to interact with $PB1_N$ are similarly conserved (Supplementary Fig. 1A, B). Anti-virals designed on the basis of $PB1_N$, therefore, may be effective against most influenza A strains. Furthermore, the high conservation of the $PB1_N$ binding site on PA suggests that anti-virals targeting this site may be less susceptible to problems of resistance associated with drugs targeting the neuraminidase and haemagglutinin.

PA_C shares similarity with part of the N-terminal domain of reovirus polymerase λ3, and we used this similarity to propose a model for the RdRp heterotrimer (Fig. 3 and Supplementary Fig. 5). The λ3 polymerase has three domains that form a hollow 'cage': the N-terminal domain I forms one side of the polymerase; the central catalytic domain II (residues 380–890) contains 'finger', 'palm' and 'thumb' subdomains for catalysis of RNA synthesis; and the C-terminal domain III covers the catalytic cleft and completes the polymerase[22]. Superimposing and substituting PA_C for the λ3 polymerase N-terminal domain shows that it is close to the λ3 catalytic domain, which has an equivalent function to PB1, with very few conflicts between the two polypeptides (Fig. 3a). PA is oriented with the mouth directed towards the λ3 catalytic domain. PA_C is predicted

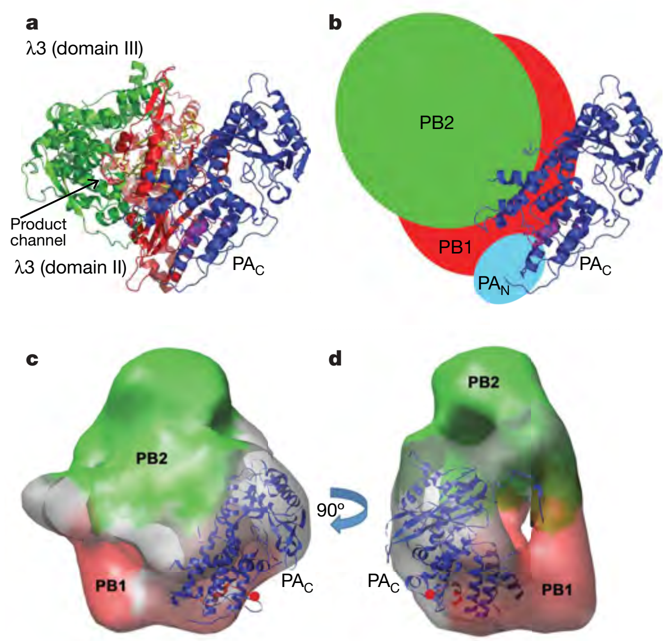


Figure 3 | Model of the influenza polymerase heterotrimer. **a**, Model of the influenza polymerase heterotrimer constructed using the λ3 reovirus polymerase structure[22]. PA_C is coloured blue, $PB1_N$ is magenta, λ3 domain II is red and the λ3 domain III is in green. **b**, Putative arrangement of PA, PB1 and PB2 in the influenza polymerase heterotrimer. Colouring is the same as in **a**, with PB1 in red and PB2 in green. **c**, **d**, PA_C docked into the electron microscopy map of the isolated heterotrimer[24]. PA_C is shown in ribbon representation and in blue. Putative locations of the PB1 and PB2 proteins are shown in red and green, respectively.

from the λ3 model to have few interactions with the C-terminal domain, consistent with observations that PA and PB2 both form stable complexes with PB1 but have no direct interaction with each other[17].

The spatial arrangement of PA, PB1 and PB2 within the heterotrimer is only partially understood. To put PA_C into the context of the larger RdRp heterotrimer, we docked the crystal structure into an available 26 Å negative staining electron microscopy map of the isolated polymerase heterotrimer (Fig. 3c)[23,24]. The isolated polymerase map shows a hollow, globular conformation and PA_C fits reasonably well into the front face of the heterotrimer, away from the site of interaction with NP monomers and in good agreement with immunolabelling using an antibody raised against residues 400–716 of PA[23,24]. Domains II and III of λ3 were loosely docked into the EM map to highlight the approximate locations of PB1 and PB2 in the heterotrimer (Fig. 3c). The EM polymerase model bears an overall resemblance to the λ3 polymerase: both form a cage-like structure with four channels providing access to the catalytic site[22].

Positioning PA_C into the polymerase heterotrimer enabled us to infer its potential roles in influenza virus replication/transcription. The conserved groove and β3–β4 loop of PA_C are located on the inner surface of the polymerase cage. The polymerase is proposed to use different binding strategies for vRNA and cRNA promoters[25], although it is not clear if it uses a single promoter binding pocket or separate binding pockets. PA is required for vRNA synthesis but preferentially binds to the cRNA promoter over the vRNA promoter, mainly by the N terminus of the protein[26]. We propose that the groove/channel in PA_C assists in recruiting cRNA; however, another study suggests that PB1 binds the cRNA promoter[27], and further work is needed to resolve this controversy.

H510 lies on the β4–β5 loop and is a potential site for endonuclease activity mediated by PB1; a H510A mutation in PA impairs nuclease activity[5]. The area including H510, which should form an endonuclease centre with PA N-terminal residues D108 and K134[15], is directed into the inner cavity of our polymerase model, where it should cleave RNA to produce primers for transport to the catalytic centre of PB1. A proposed nucleotide binding motif from residues 502–509 resembles the GXXXXGKT/S motif[28] and lies on the same loop as H510 (Supplementary Fig. 1A, yellow box). The conserved D529 site described previously[28] is too far to co-operate with this proposed NTP-binding motif, although D478 on the α4–β4 turn is close to the motif and in a region that interacts with R508 and H510. Another possible role for the RNA groove in PA_C may be for host messenger RNA binding before endonuclease cleavage, which is consistent with the endonuclease defect caused by mutation of H510, although this remains to be confirmed by further studies. Because the N-terminal one-third of PA is also reported to be required for polymerase endonuclease activity[15], it should be located adjacent to the region including H510 of PA_C. Further studies are required to determine the precise location and function of the PA N terminus in the heterotrimer.

We report the crystal structure of a principal part of the influenza polymerase PA protein in complex with an inhibitory peptide of PB1. The availability of the PA_C–$PB1_N$ structure offers a starting point for further investigation into the structure and function of the influenza virus polymerase. Together with recent structures of PB2 domains for nuclear import[29] and cap binding[4], the PA_C–$PB1_N$ structure reported here will increase our understanding of the molecular basis for influenza virus replication and transcription. Moreover, there is an urgent need for improved influenza therapeutics; the PA_C–$PB1_N$ complex structure provides a key conserved target for the design of a new generation of compounds that inhibit polymerase assembly and activity.

METHODS SUMMARY

Residues 257–716 of the avian H5N1 influenza A virus (A/goose/Guangdong/1/96) PA gene were cloned into the pGEX-6p vector (GE Healthcare) and

overexpressed in *Escherichia coli* strain BL21. A gene fragment encoding residues 1–25 of PB1 was also cloned into the same vector and overexpressed in BL21 cells. Cell cultures expressing each of these two proteins were mixed in approximately 1:1 molar ratios of expressed proteins. Recombinant proteins were co-purified with a glutathione affinity column (GE Healthcare). Glutathione *S*-transferase (GST) was cleaved with PreScission protease (GE Healthcare), and the protein complex was further purified by Q ion exchange chromatography and Superdex-200 gel filtration chromatography (GE Healthcare).

The PA_C–$PB1_N$ complex was crystallized in the space group $P4_12_12$ using 1–1.5 M sodium acetate as the precipitant at a pH of 7.9. The structure was phased to 3.6 Å by multiple-wavelength anomalous dispersion from a selenomethionyl derivative, and traced using 2.9 Å native data. The final refined model, with an R factor of 22% and an R_{free} of 26%, contains residues 257–716 of PA (with some missing internal fragments) and residues 1–15 of PB1.

Double mutations in PA_C were introduced using the PCR method by designing mutated residues in primers. The mutated genes were also cloned into the pGEX-6p vector. For *in vitro* binding experiments, the wild-type or double mutant PA_C proteins were separately purified from *E. coli* strain BL21. Purified GST–$PB1_N$ fusion peptide or GST alone (native control) were first immobilized on glutathione affinity-column resins (GSH), followed by addition of approximately fivefold molar excess of purified wild-type/double mutant PA_C. GSH resins were extensively washed with PBS (pH 7.4) to remove unbound proteins. The bound protein was eluted from GSH resins with GST elution buffer containing 10 mM reduced glutathione. The eluted sample was analysed by SDS–PAGE electrophoresis and stained by Coomassie blue.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 March; accepted 23 May 2008.
Published online 9 July 2008.

1. Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889–893 (2005).
2. Ghanem, A. *et al.* Peptide-mediated interference with influenza A virus polymerase. *J. Virol.* **81**, 7801–7804 (2007).
3. Li, M. L., Rao, P. & Krug, R. M. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *EMBO J.* **20**, 2078–2086 (2001).
4. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Biol.* **15**, 500–506 (2008).
5. Fodor, E. *et al.* A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J. Virol.* **76**, 8989–9001 (2002).
6. Jung, T. E. & Brownlee, G. G. A new promoter-binding site in the PB1 subunit of the influenza A virus polymerase. *J. Gen. Virol.* **87**, 679–688 (2006).
7. Kawaguchi, A., Naito, T. & Nagata, K. Involvement of influenza virus PA subunit in assembly of functional RNA polymerase complexes. *J. Virol.* **79**, 732–744 (2005).
8. Huarte, M. *et al.* Threonine 157 of influenza virus PA polymerase subunit modulates RNA replication in infectious viruses. *J. Virol.* **77**, 6007–6013 (2003).
9. Fodor, E., Mingay, L. J., Crow, M., Deng, T. & Brownlee, G. G. A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase promotes the generation of defective interfering RNAs. *J. Virol.* **77**, 5017–5020 (2003).
10. Sugiura, A., Ueda, M., Tobita, K. & Enomoto, C. Further isolation and characterization of temperature-sensitive mutants of influenza virus. *Virology* **65**, 363–373 (1975).
11. Zurcher, T., de la Luna, S., Sanz-Ezquerro, J. J., Nieto, A. & Ortin, J. Mutational analysis of the influenza virus A/Victoria/3/75 PA protein: studies of interaction with PB1 protein and identification of a dominant negative mutant. *J. Gen. Virol.* **77**, 1745–1749 (1996).
12. Hara, K. *et al.* Influenza virus RNA polymerase PA subunit is a novel serine protease with Ser624 at the active site. *Genes Cells* **6**, 87–97 (2001).
13. Rodriguez, A., Perez-Gonzalez, A. & Nieto, A. Influenza virus infection causes specific degradation of the largest subunit of cellular RNA polymerase II. *J. Virol.* **81**, 5315–5324 (2007).
14. Sanz-Ezquerro, J. J., Zurcher, T., de la Luna, S., Ortin, J. & Nieto, A. The amino-terminal one-third of the influenza virus PA protein is responsible for the induction of proteolysis. *J. Virol.* **70**, 1905–1911 (1996).
15. Hara, K., Schmidt, F. I., Crow, M. & Brownlee, G. G. Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. *J. Virol.* **80**, 7789–7798 (2006).
16. Regan, J. F., Liang, Y. & Parslow, T. G. Defective assembly of influenza A virus due to a mutation in the polymerase subunit PA. *J. Virol.* **80**, 252–261 (2006).
17. Toyoda, T., Adyshev, D. M., Kobayashi, M., Iwata, A. & Ishihama, A. Molecular assembly of the influenza virus RNA polymerase: determination of the subunit–subunit contact sites. *J. Gen. Virol.* **77**, 2149–2157 (1996).
18. Ohtsu, Y., Honda, Y., Sakata, Y., Kato, H. & Toyoda, T. Fine mapping of the subunit binding sites of influenza virus RNA polymerase. *Microbiol. Immunol.* **46**, 167–175 (2002).
19. Perez, D. R. & Donis, R. O. Functional analysis of PA binding by influenza a virus PB1: effects on polymerase activity and viral infectivity. *J. Virol.* **75**, 8127–8136 (2001).
20. Gonzalez, S., Zurcher, T. & Ortin, J. Identification of two separate domains in the influenza virus PB1 protein involved in the interaction with the PB2 and PA subunits: a model for the viral RNA polymerase structure. *Nucleic Acids Res.* **24**, 4456–4463 (1996).
21. Gabriel, G. *et al.* Differential polymerase activity in avian and mammalian cells determines host range of influenza virus. *J. Virol.* **81**, 9601–9604 (2007).
22. Tao, Y., Farsetta, D. L., Nibert, M. L. & Harrison, S. C. RNA synthesis in a cage–structural studies of reovirus polymerase λ3. *Cell* **111**, 733–745 (2002).
23. Area, E. *et al.* 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proc. Natl Acad. Sci. USA* **101**, 308–313 (2004).
24. Torreira, E. *et al.* Three-dimensional model for the isolated recombinant influenza virus polymerase heterotrimer. *Nucleic Acids Res.* **35**, 3774–3783 (2007).
25. Deng, T., Vreede, F. T. & Brownlee, G. G. Different de novo initiation strategies are used by influenza virus RNA polymerase on its cRNA and viral RNA promoters during viral RNA replication. *J. Virol.* **80**, 2337–2348 (2006).
26. Maier, H. J., Kashiwagi, T., Hara, K. & Brownlee, G. G. Differential role of the influenza A virus polymerase PA subunit for vRNA and cRNA promoter binding. *Virology* **370**, 194–204 (2008).
27. Gonzalez, S. & Ortin, J. Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. *EMBO J.* **18**, 3767–3775 (1999).
28. de la Luna, S., Martinez, C. & Ortin, J. Molecular cloning and sequencing of influenza virus A/Victoria/3/75 polymerase genes: sequence evolution and prediction of possible functional domains. *Virus Res.* **13**, 143–155 (1989).
29. Tarendeau, F. *et al.* Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nature Struct. Biol.* **14**, 229–233 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Chen and K. Yu for providing the A/goose/Guangdong/1/96 influenza PA gene; J. Ortin and O. Llorca for supplying electron microscopy maps; C. Yang, X. Su, F. Sun, L. Wang and R.-M. Xu for advice and technical assistance; and S. Harrison, P. Kuhn, X. Chen and T. Toyoda for discussion. This work was supported by the National Natural Science Foundation of China (grant numbers 30599432 and 30221003), the Ministry of Science and Technology International Cooperation Project (grant number 2006DFB32420), the Ministry of Science and Technology 863 Project (grant numbers 2006AA02A314 and 2006AA02A322) and the Ministry of Science and Technology 973 Project (grant numbers 2006CB504300 and 2007CB914300).

Author Information Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data Bank under accession number 3CM8. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Z.R. (raozh@xtal.tsinghua.edu.cn) and Y.L. (liuy@ibp.ac.cn).

METHODS

Protein expression and purification. Residues 257–716 of the avian H5N1 influenza A virus (A/goose/Guangdong/1/96) PA gene were cloned into the pGEX-6p vector (GE Healthcare) and overexpressed in *E. coli* strain BL21. A gene fragment that encodes the N-terminal 25 residues of PB1 was also cloned into the same vector and overexpressed in BL21 cells. Cell cultures expressing each of these two proteins were mixed in approximately 1:1 molar ratios of expressed proteins. The recombinant proteins were then co-purified with a glutathione affinity column (GE Healthcare). After cleavage of the GST tag with PreScission protease (GE Healthcare), the protein complex was further purified by Q ion exchange chromatography and then by Superdex-200 gel filtration chromatography (GE Healthcare).

***In vitro* binding experiments.** Double point mutations in PA_C were introduced using the PCR method by designing mutated residues in primers. The mutated genes were also cloned into the pGEX-6p vector. For *in vitro* binding experiments, the wild-type or double mutant PA_C proteins were separately purified from *E. coli* strain BL21 as described in the main text. The purified GST–$PB1_N$ fusion peptide or GST alone (native control) was first immobilized on GSH resins, followed by the addition of approximately fivefold molar excess of purified wild-type/double mutant PA_C. GSH resins were extensively washed with PBS (pH 7.4) to remove unbound proteins, and the bound protein was eluted from the GSH resins with GST elution buffer containing 10 mM reduced glutathione. The eluted sample was then analysed by SDS–PAGE electrophoresis and stained by Coomassie blue.

Crystallization and structure determination. Crystals were obtained by the vapour diffusion method with 1–1.5 M sodium acetate as the precipitant at pH 7.9. Native 2.9 Å data were collected in-house ($\lambda = 1.5418$ Å) using an FR-E SuperBright rotating anode X-ray source equipped with an R-AXIS IV++ image plate detector (Rigaku). Anomalous X-ray diffraction data were collected to 3.6 Å at peak ($\lambda = 0.9783$ Å) and inflection ($\lambda = 0.9785$ Å) wavelengths from a selenomethionyl-derivative protein at the Advanced Photon Source, Chicago. Diffraction data were processed by HKL-2000 (ref. 30). Peak and inflection data sets were used for phasing by the multi-wavelength anomalous dispersion method[31]. A total of 17 expected heavy atoms were located by SHELXD[32] and initial phases were calculated using SHARP[33]. Density modification was performed by DM[34] and Solomon[35]. Some secondary structure elements were clearly observed in the experimental electron density map. Approximately 60% of residues were traced automatically by ARP/wARP[36] after phase extension by Phenix[37] using 2.9 Å native data and phases to 3.6 Å. The remainder of the model was built manually using the program Coot[38]. The model was refined against native data at 2.9 Å resolution using the programs CNS[39] and Refmac[40]. The final refined model, with an *R* factor of 22% and an R_{free} of 26%, contains residues 257–716 of PA with some missing internal fragments, and residues 1–15 of PB1. All diagrams were prepared using PyMOL (http://www.pymol.org) and electrostatic surface charges were calculated by APBS. Docking of the PA_C crystal structure into the low-resolution electron microscopy map and visualization of the electron microscopy model were performed with UCSF Chimera (http://www.cgl.ucsf.edu/chimera/).

30. Otwinowski, Z. & Minor, W. in *Macromolecular Crystallography*, part A (eds Carter C. W. Jr & Sweet, R. M.) 307–326 (Academic, New York, 1997).
31. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254,** 51–58 (1991).
32. Sheldrick, G. M. *Direct Methods for Solving Macromolecular Structures* (ed. Fortier, S.) 401–411 (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998).
33. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364,** 215–230 (2007).
34. Cowtan, K. D. & Zhang, K. Y. Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72,** 245–270 (1999).
35. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52,** 30–42 (1996).
36. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6,** 458–463 (1999).
37. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
39. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
40. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).

 nature

LETTERS

The structural basis for an essential subunit interaction in influenza virus RNA polymerase

Eiji Obayashi[1], Hisashi Yoshida[1], Fumihiro Kawai[1], Naoya Shibayama[2], Atsushi Kawaguchi[3], Kyosuke Nagata[3], Jeremy R. H. Tame[1] & Sam-Yong Park[1]

Influenza A virus is a major human and animal pathogen with the potential to cause catastrophic loss of life. The virus reproduces rapidly, mutates frequently and occasionally crosses species barriers. The recent emergence in Asia of avian influenza related to highly pathogenic forms of the human virus has highlighted the urgent need for new effective treatments[1]. Here we demonstrate the importance to viral replication of a subunit interface in the viral RNA polymerase, thereby providing a new set of potential drug binding sites entirely independent of surface antigen type. No current medication targets this heterotrimeric polymerase complex. All three subunits, PB1, PB2 and PA, are required for both transcription and replication[2–4]. PB1 carries the polymerase active site, PB2 includes the capped-RNA recognition domain, and PA is involved in assembly of the functional complex[5–7], but so far very little structural information has been reported for any of them[8–11]. We describe the crystal structure of a large fragment of one subunit (PA) of influenza A RNA polymerase bound to a fragment of another subunit (PB1). The carboxy-terminal domain of PA forms a novel fold, and forms a deep, highly hydrophobic groove into which the amino-terminal residues of PB1 can fit by forming a 3_{10} helix.

Influenza A virus carries eight negative-strand RNA genome segments, each bound by the virally encoded RNA-dependent RNA polymerase complex, which has a molecular mass of around 250 kDa[12]. This complex carries out numerous essential roles in viral replication and pathogenesis, including 'cap-snatching'; that is, the cleavage of host cell pre-messenger RNA to utilize its cap for viral transcripts[7,13]. The precise roles of each subunit are still being actively investigated. Because the complex is essential for many processes it shows a high level of sequence conservation across strains[14]. It is also unrelated to human proteins, and therefore makes an appealing drug target. Currently however, there is very limited structural information for the complex. A 23 Å resolution reconstruction of the overall structure by electron microscopy shows a compact shape with no obvious domain boundaries[8,9]. The C-terminal domain of PB2 including its nuclear localization signal has been crystallized, in complex with importin α5 (ref. 10). More recently a central domain from the same subunit has been crystallized and shown to carry the cap-binding site[11]. The subunit interactions in RNA polymerase have been characterized by extensive mutagenesis, showing that the N-terminal tip of PB1 binds to the C terminus of PA[15–18], and the loss of PA abolishes RNA polymerase activity and viral replication[6]. PA is involved in the assembly of functional polymerase complex, cap binding and virion RNA (vRNA) promoter binding[6,19]. PA and its interface with PB1 are therefore potential drug targets that are currently not being exploited.

Tryptic digestion shows that PA forms two domains[19]. The C-terminal domain carrying the PB1 binding site (residues 239–716) can be overexpressed alone in *Escherichia coli* and purified. Pull-down assays were used to confirm the reports[15–18] that the C-terminal region (residues 657–716) is required for binding the N terminus of PB1 (see Supplementary Information). Co-expression of PA (residues 239–716) with PB1 (residues 1–81) yielded a stable complex which could be purified. Similar results were obtained when PB1 was truncated to the N-terminal 14 or 36 residues, but only PA(239–716)–PB1(1–81) yielded diffracting crystals.

Native X-ray diffraction data were collected to 2.3 Å, and the structure was solved using a single mercury-soaked crystal, revealing that the C domain of PA consists of 13 α-helices and 9 β-strands (Fig. 1a, b). Eighteen residues at the N terminus and a loop (residues 372–397) between α3 and α4 are disordered, giving a total of 423 residues of PA in the final model, beginning with Ile 257. Overall 15 residues of PB1 are visible in the electron density map, from Met 1 to Gln 15 (Supplementary Fig. 1). These residues are completely conserved in human and avian influenza (Fig. 1c), and form extensive contacts with PA (Fig. 2 and 3). Protein Data Bank (PDB) searches revealed no similar structure to PA. Three α-helices, α10, α11 and α13, are positioned like the jaws of a clamp, grasping the N terminus of PB1 with the support of a β-hairpin loop made by β8 and β9 (Fig. 1b and Supplementary Fig. 2a). Met 1 and Asp 2 of PB1 emerge from a gap beside the hairpin loop and are highly solvent exposed, whereas the side chain of Val 3 is partly buried (Fig. 2). Even if *N*-acetylated, Met 1 seems unlikely to form strong interactions with PA. Pro 5 to Lys 11 form a 3_{10} helix, which is held by the finger-like clasp of PA. Pro 13 turns the main chain so that Ala 14 and Gln 15 also pack against PA, but no further residues appear ordered in the structure. Mass spectrometry confirmed that all 81 residues of the PB1-derived peptide are present in the crystal, but most are invisible in the electron density.

PB1 interacts with PA through an array of hydrogen bonds and hydrophobic contacts (Fig. 2 and 3). Most inter-subunit hydrogen bonds form through main-chain atoms of PB1. Residues Asp 2 to Asn 4 form anti-parallel β-sheet-like interactions with Ile 621 to Glu 623 of PA. The carbonyl oxygen atoms of Asp 2, Val 3, Phe 9, Leu 10 and Val 12 in PB1 form hydrogen bonds to Glu 623, Gln 408, Trp 706, Gln 670 and Arg 673 in PA, and the backbone nitrogen atoms of Asp 2, Val 3, Asn 4, Leu 8, and Ala 14 in PB1 form hydrogen bonds with Glu 623, Asn 412, Ile 621, Pro 620 and Gln 670, respectively (Fig. 2a). Hydrophobic interactions seem to contribute substantially to the binding energy (Fig. 3). Pro 5 packs between Phe 411 and Trp 706, and Leu 8 makes contact with the side chains of Met 595, Trp 619, Val 636 and Leu 640. Using this model, we designed deletions and point mutations in the C-terminal domain

[1]Protein Design Laboratory, Yokohama City University, 1-7-29 Suehiro, Tsurumi, Yokohama 230-0045, Japan. [2]Department of Physiology, Division of Biophysics, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke, Tochigi 329-0498, Japan. [3]Department of Infection Biology, Graduate School of Comprehensive Human Sciences and Institute of Basic Medical Sciences, University of Tsukuba, Tsukuba 305-8575, Japan.

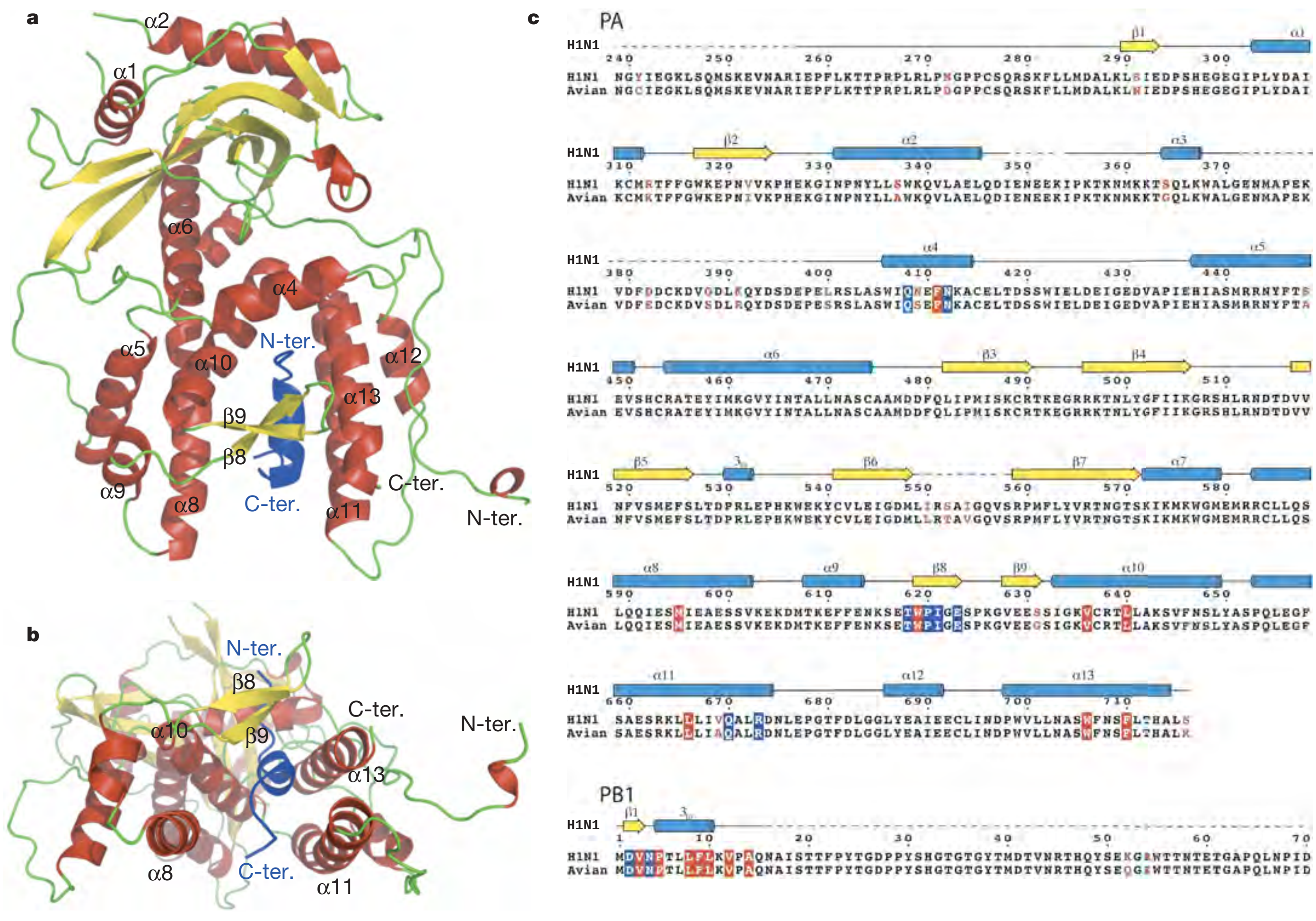


Figure 1 | Crystal structure of the C-terminal domain of PA bound to the N-terminal peptide of PB1. a, An overall ribbon diagram showing the fold of PA, with helices coloured red, strands yellow and coil green. Helices are numbered from the N terminus. PB1 residues are coloured dark blue. The principal β-sheet is formed largely from residues 482–571. **b**, The same model as **a** but rotated 90° around a horizontal axis to show a view down the 'clamp' made from α-helices 10, 11 and 13. **c**, The sequence alignment of human (H1N1) influenza PA with that from an avian strain (AY059532 for A/Duck/HK/00, H5N1). Non-conserved residues are shown in red. Secondary structure is indicated with blue bars showing helices, yellow arrows showing strands, and broken lines showing disordered regions. Amino acid residues shown in white on blue form hydrogen bonds across the PA–PB1 interface; residues shown in white on red form interface hydrophobic contacts.

of PA that greatly weaken or abolish PB1 binding, and similarly reduce viral RNA synthesis in human cells (Fig. 4). The levels of vRNA, complementary RNA (cRNA) and viral mRNA synthesis were markedly lowered for all the mutants (Fig. 4c).

The interactions observed in the model are completely consistent with previous aspartate scanning mutagenic studies that showed that only Pro 5, Leu 7, Leu 8, Phe 9 and Leu 10 appear essential for binding to PA[16]. Mutation of Pro 5 of PB1 to leucine abolished binding, suggesting that this residue helps to stabilize the helix as well as provide apolar contacts with PA. In contrast, replacing either Val 3 or Thr 6 with aspartate only reduced binding (by about 80% and 25%, respectively, in the assay used), because these side chains are accessible to solvent water in the complex. Both Leu 7 and Leu 8 form many hydrophobic contacts and their replacement with a charged side chain would be expected to destabilize the complex strongly. Mutating either Phe 9 or Leu 10 to aspartate also prevented PB1 binding completely, even though these residues do not apparently form as many interactions with PA as Leu 7 and Leu 8, and are solvent exposed. Possibly an aspartate side chain at position 9 of PB1 would disrupt the helix of the N-terminal peptide. Equally, a carboxyl group in this position might distort the binding pocket by pulling on Lys 643 or Arg 663 of PA, which lie nearby. Leu 10 of PB1 contacts the side chain of Leu 7, and its replacement with aspartate may interfere with key interactions formed by that residue. Asp 10 of the mutant would almost certainly be pulled away from Leu 7 by Arg 673 in PA. Overall the crystal structure provides a clear explanation for the behaviour of the PB1 mutants, and strongly suggests that the core of the PB1 interaction interface is restricted to five residues: Pro 5, Leu 7, Leu 8, Phe 9 and Leu 10. The contribution of Phe 9 may in fact be smaller, and a leucine or similar residue in this position might be capable of supporting strong PB1–PA binding. Only the last two residues of the PTLLFL peptide sequence form hydrogen bonds with PA.

A number of genetic and biochemical experiments have indicated residues in PA responsible for some activities of the viral polymerase (Supplementary Fig. 2). Lys 102 and His 510 are required for the cap-binding activity[19,20]. Thr 157, Leu 226, Arg 442, Arg 443, Glu 493, Gly 494, His 510, Glu 524, Lys 536, Trp 537, Glu 656 and Gly 657 are implicated in the replication activity[6,20–22]. Mapping these residues on the model of PA shows them to be widely distributed throughout the protein. His 510 and Glu 656 are external, but many other side chains from this group form salt bridges or polar interactions which stabilize the protein fold. For example, Glu 524 forms a deeply buried hydrogen bond with the main-chain nitrogen atom of Leu 283. Mutation of Cys 453 or Arg 638 has been shown to cause synthesis of defective interfering RNA[5], and these two residues lie in van der Waals contact in the crystal structure. The cysteine is buried beneath the arginine, which makes a salt bridge with Glu 449.

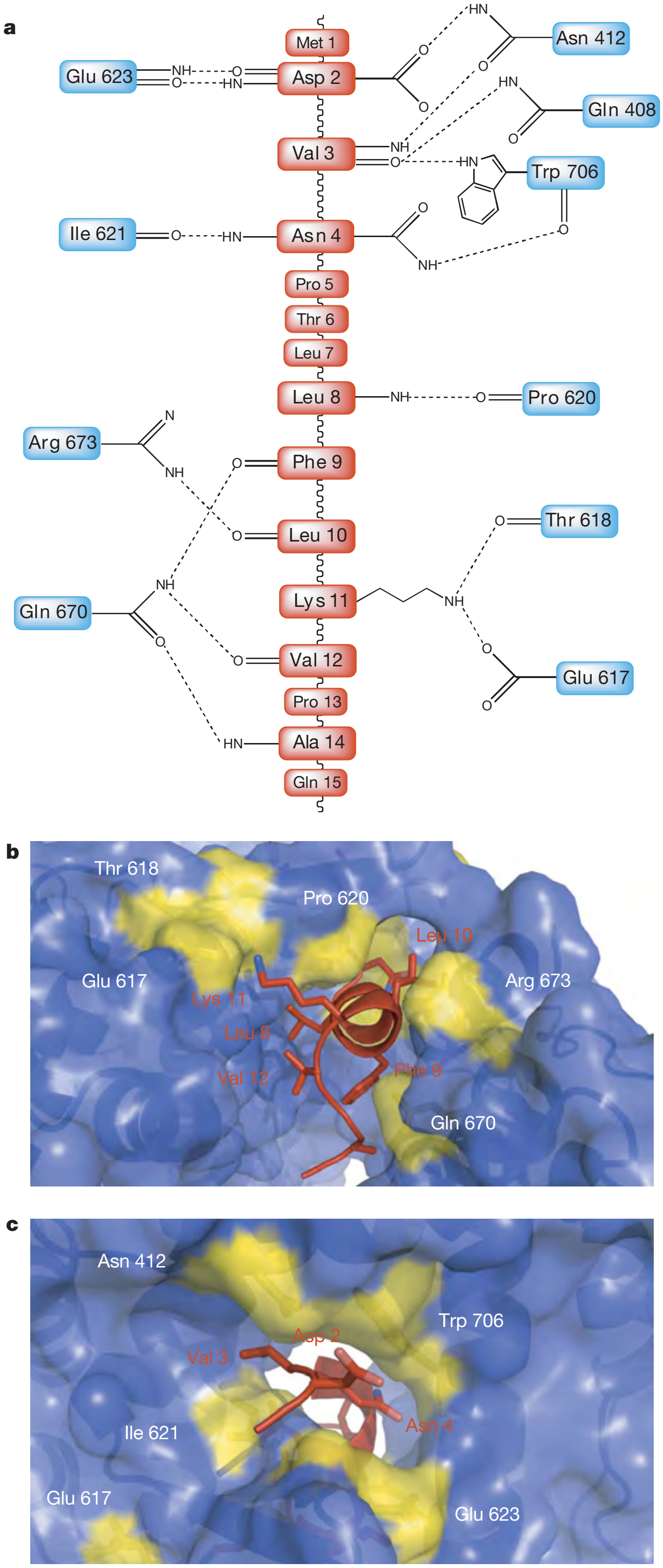


Figure 2 | Hydrogen bonds between PA and PB1. **a**, Schematic diagram showing the hydrogen bonds between PA (blue boxes) and PB1 (orange boxes). Black dashed lines indicate hydrogen bonds between 2.4–3.4 Å in length. **b**, Molecular surface representation showing the cleft into which PB1 binds. The surface of PA is coloured blue, except residues that form hydrogen bonds with PB1, which are shown in yellow. PB1 is shown in orange Cα trace, with labelled residues shown as sticks, nitrogen atoms coloured blue and oxygen atoms pink. **c**, The same model as **b** but rotated by 180° about a horizontal axis to show the N terminus of PB1.

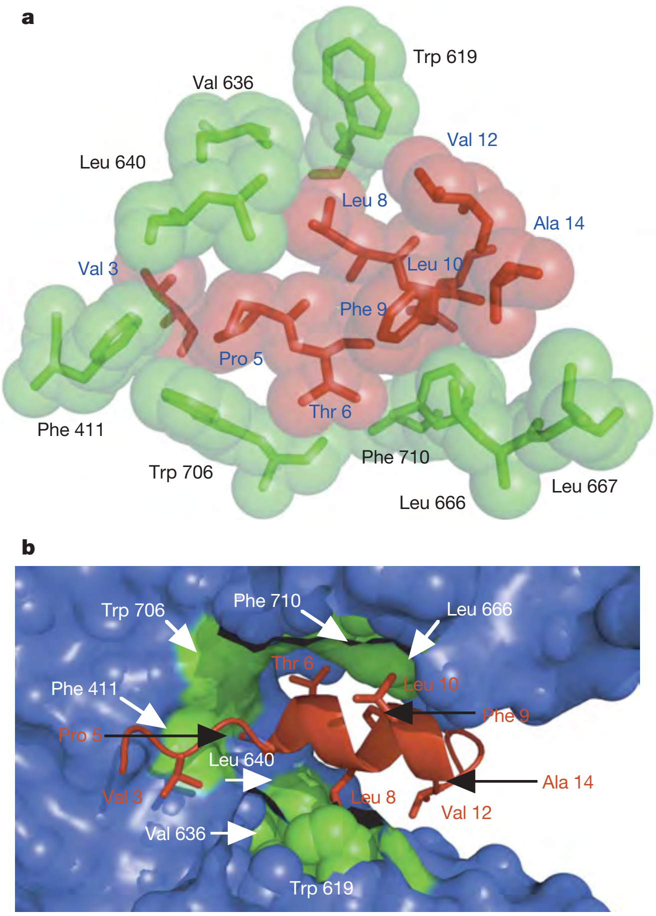


Figure 3 | Hydrophobic contacts between PA and PB1. **a**, Space-filling representation, with PA residues shown in green and PB1 residues in orange. Although the interface is largely tightly packed, unfilled space is found adjacent to Pro 5, Thr 6 and Phe 9 of PB1. The absence of any contact between Thr 6 of PB1 and PA suggests that substitutions at this position could exploit interactions with the nearby side chains of Leu 666 and Phe 710. Leu 666 is pressed against the benzene ring of Phe 9, with a distance between the side chains of 3.6 Å. **b**, A cut-away diagram showing the binding site, with β8 and β9 removed. The molecular surface of PA is coloured green for residues making hydrophobic contacts (between 3.5–4.3 Å in length) with PB1, and blue for all other residues. PB1 is shown in orange.

Most current influenza drugs target either haemagglutinin (HA) or neuraminidase (NA), two major antigens present at the virion surface[23]. Sixteen different HA subtypes and nine different NA subtypes have been identified[24]. Oseltamivir (sold as Tamiflu) and zanamivir (Relenza), for example, are NA inhibitors, and prevent viral particles being released from infected cells[25–28]. Billions of pounds worth of oseltamivir were stockpiled worldwide in response to the last Asian epidemic, but resistant influenza is already emerging. The anti-influenza drug amantadine targets the M2 protein, a viral proton channel[29,30]; however, a single residue change is sufficient to confer resistance, which has risen sufficiently to render the drug useless against many strains. Both oseltamivir and amantadine target proteins with a single known function and substantial sequence variation between viral strains. There is therefore ample scope to develop new lead molecules disrupting other processes in the viral life cycle. The highly conserved PB1 binding site on PA may have considerable potential as a drug target site, given that the interaction is crucial to many viral functions and rests on a handful of hydrophobic groups and hydrogen bonds. The peptide PTLLFL is clearly a lead molecule that might assist the development of new treatments effective against all types of influenza A virus, including avian strains, and it is hoped that the structure presented here will initiate this process.

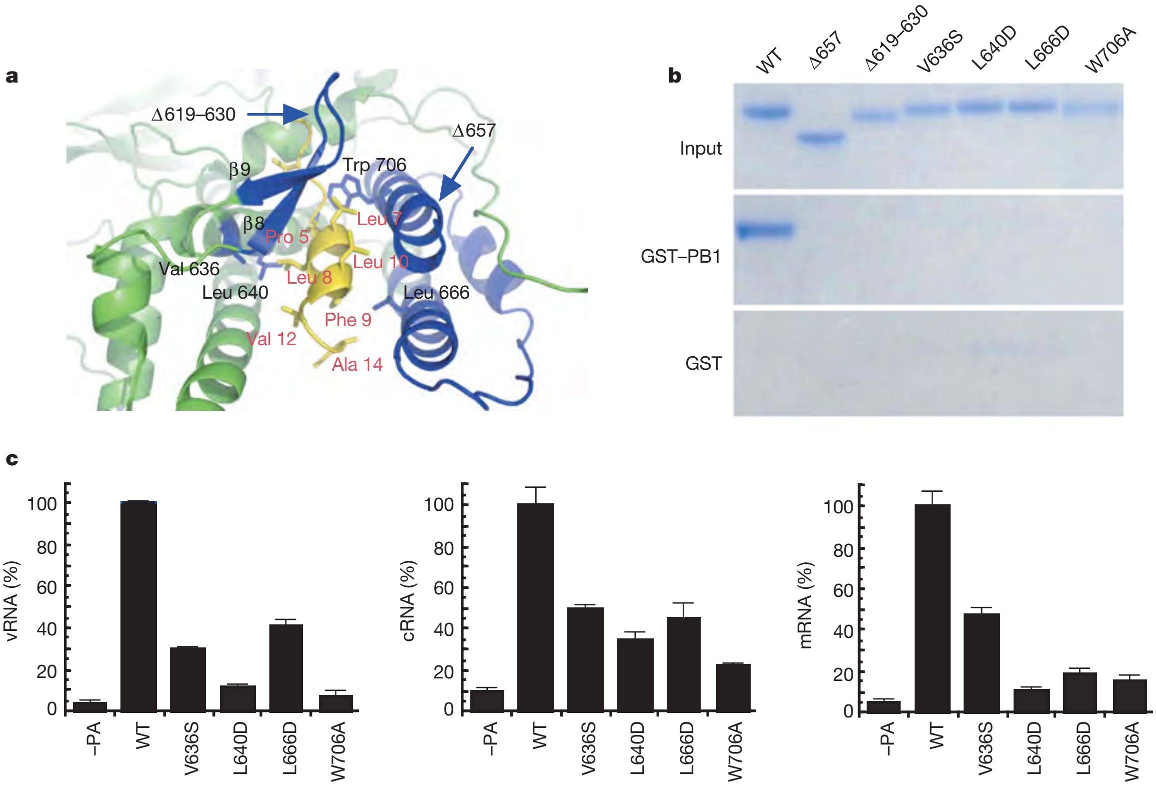


Figure 4 | Transcriptional activity of PA mutants. a, Cartoon showing the Cα trace of PA in green, with residues selected for mutagenesis or deletion shown in blue. PB1 is shown in yellow (residues labelled in red). Val 636 touches Leu 8, Leu 640 lies close to Leu 8 and Pro 5, Leu 666 packs against the side chain of Phe 9, and Trp 706 interacts with Asn 4, Pro 5 and Thr 6. **b,** GST pull-down assay. Wild-type (WT) PA and various mutants were tested for binding to GST fused to the N-terminal 14 residues of PB1 (middle) or by GST alone as a negative control (bottom). The eluted proteins were analysed by 5–20% SDS–PAGE and Coomassie blue staining, showing that each mutation greatly weakened the interaction *in vitro*. **c,** Effects of various mutations of PA on the level of viral genomic RNA (vRNA), complementary RNA (cRNA) and viral mRNA in reporter assays (see Methods). In the absence of PA, the levels of each RNA fall very low. The single site mutations also exhibited large effects, generally reducing the yield of each RNA by at least 40%. Results are means and s.d. for three independent experiments.

METHODS SUMMARY

Cloning, expression, purification of PA–PB1 complex and mutants assay. The genes encoding PA and PB1 from influenza A/Puerto Rico/8/1934 H1N1 were cloned into pET28b with a hexa-histidine tag fused at the N terminus of PA and were co-expressed in *E. coli* strain BL21-CodonPlus(DE3)-RILP (Stratagene). The expressed complex was purified by Ni-NTA (Qiagen) and Q-sepharose (GE Healthcare). For pull-down assays, PA mutants were cloned and purified as complexes with the PB1 fragment by the same method. A gene encoding PB1 residues 1–14 was cloned into pET28b with a glutathione *S*-transferase (GST) tag fused at its N terminus. After incubation of wild-type and mutant PAs with GST-fused PB1, proteins were pulled down by glutathione sepharose (GE Healthcare) and analysed by SDS-acrylamide gel electrophoresis.

Crystallization, structure determination and refinement. Crystals grew in space group $P3_221$, with $a = b = 101.9$ Å, $c = 115.0$ Å, and contained one molecule in an asymmetric unit. The final *R* and *R*-free factors are 20.7% and 26.2% at 2.3 Å resolution, respectively. 87.7% of the residues in the final model are in the most favourable regions of the Ramachandran plot, with no residues in disallowed regions (Supplementary Table 1).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 April; accepted 2 July 2008.
Published online 27 July 2008.

1. Peiris, J. S., de Jong, M. D. & Guan, Y. Avian influenza virus (H5N1): a threat to human health. *Clin. Microbiol. Rev.* **20,** 243–267 (2007).
2. Horisberger, M. A. The large P proteins of influenza A viruses are composed of one acidic and two basic polypeptides. *Virology* **107,** 302–305 (1980).
3. Braam, J., Ulmanen, I. & Krug, R. M. Molecular model of a eucaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription. *Cell* **34,** 611–618 (1983).
4. Nagata, K., Kawaguchi, A. & Naito, T. Host factors for replication and transcription of the influenza virus genome. *Rev. Med. Virol.* **18,** 247–260 (2008).
5. Fodor, E., Mingay, L. J., Crow, M., Deng, T. & Brownlee, G. G. A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase promotes the generation of defective interfering RNAs. *J. Virol.* **77,** 5017–5020 (2003).
6. Kawaguchi, A., Naito, T. & Nagata, K. Involvement of influenza virus PA subunit in assembly of functional RNA polymerase complexes. *J. Virol.* **79,** 732–744 (2005).
7. Deng, T., Sharps, J. L. & Brownlee, G. G. Role of the influenza virus heterotrimeric RNA polymerase complex in the initiation of replication. *J. Gen. Virol.* **87,** 3373–3377 (2006).
8. Area, E. *et al.* 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proc. Natl Acad. Sci. USA* **101,** 308–313 (2004).
9. Torreira, E. *et al.* Three-dimensional model for the isolated recombinant influenza virus polymerase heterotrimer. *Nucleic Acids Res.* **35,** 3774–3783 (2007).
10. Tarendeau, F. *et al.* Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nature Struct. Mol. Biol.* **14,** 229–233 (2007).
11. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Mol. Biol.* **15,** 500–506 (2008).
12. Elton, D., Digard, P., Tiley, L. & Ortin, J. in *Current Topics in Influenza Virology* (ed. Kawaoka, Y.) 1–92 (Horizon Scientific Press, 2005).
13. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m^{7}GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23,** 847–858 (1981).
14. Taubenberger, J. K. *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature* **437,** 889–893 (2005).
15. Zurcher, T., de la Luna, S., Sanz-Ezquerro, J. J., Nieto, A. & Ortin, J. Mutational analysis of the influenza virus A/Victoria/3/75 PA protein: studies of interaction with PB1 protein and identification of a dominant negative mutant. *J. Gen. Virol.* **77,** 1745–1749 (1996).
16. Perez, D. R. & Donis, R. O. Functional analysis of PA binding by influenza A virus PB1: effects on polymerase activity and viral infectivity. *J. Virol.* **75,** 8127–8136 (2001).
17. Ohtsu, Y., Honda, Y., Sakata, Y., Kato, H. & Toyoda, T. Fine mapping of the subunit binding sites of influenza virus RNA polymerase. *Microbiol. Immunol.* **46,** 167–175 (2002).
18. Ghanem, A. *et al.* Peptide-mediated interference with influenza A virus polymerase. *J. Virol.* **81,** 7801–7804 (2007).
19. Hara, K., Schmidt, F. I., Crow, M. & Brownlee, G. G. Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a

critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. *J. Virol.* **80,** 7789–7798 (2006).
20. Fodor, E. *et al.* A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J. Virol.* **76,** 8989–9001 (2002).
21. Huarte, M. *et al.* Threonine 157 of influenza virus PA polymerase subunit modulates RNA replication in infectious viruses. *J. Virol.* **77,** 6007–6013 (2003).
22. Regan, J. F., Liang, Y. & Parslow, T. G. Defective assembly of influenza A virus due to a mutation in the polymerase subunit PA. *J. Virol.* **80,** 252–261 (2006).
23. Hsieh, H. P. & Hsu, J. T. Strategies of development of antiviral agents directed against influenza virus replication. *Curr. Pharm. Des.* **13,** 3531–3542 (2007).
24. World Health Organization.. A revision of the system of nomenclature for influenza viruses: a WHO memorandum. *Bull. World Health Organ.* **58,** 585–591 (1980).
25. Kim, C. U. *et al.* Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: Design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* **119,** 681–690 (1997).
26. von Itzstein, M. *et al.* Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363,** 418–423 (1993).
27. Russell, R. J. *et al.* The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* **443,** 45–49 (2006).
28. Liu, Y., Zhang, J. & Xu, W. Recent progress in rational drug design of neuraminidase inhibitors. *Curr. Med. Chem.* **14,** 2872–2891 (2007).
29. Wang, C., Takeuchi, K., Pinto, L. H. & Lamb, R. A. Ion channel activity of influenza A virus M2 protein: characterization of the amantadine block. *J. Virol.* **67,** 5585–5594 (1993).
30. Stouffer, A. L. *et al.* Structural basis for the function and inhibition of an influenza virus proton channel. *Nature* **451,** 596–599 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff at the SPring-8 BL41XU and Photon Factory BL5A beamlines for assistance with data collection, and M. Yao and Y. Zhou for help with data processing. This work was supported in part by the ISS applied research partnership program (S.-Y.P.).

Author Contributions E.O., J.R.H.T., K.N. and S.-Y.P. conceived and designed the project. E.O. carried out complex purification and crystallization, and pull-down assays. E.O., H.Y., F.K., N.S. and S.-Y.P. conducted experimental work including cloning, expression, purification, binding assays, data collection and structure analysis. A.K. and K.N. tested the transcriptional activity of mutants. E.O., J.R.H.T. and S.-Y.P. wrote the manuscript, and all authors discussed the results and conclusions.

Author Information Atomic coordinates and structure factors of the complex have been deposited in the Protein Data Bank under accession code 2ZNL. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.-Y.P. (park@tsurumi.yokohama-cu.ac.jp).

METHODS

Cloning, expression and purification of the PA–PB1 complex. The genes encoding PA and PB1 were amplified by PCR using cDNA (influenza A/Puerto Rico/8/1934, H1N1) in pET14b (ref. 31). The product encoding PA residues 239–716 was digested with BamHI and NotI and then ligated into suitably cut, modified pET28b vector, in which a SD sequence, initial ATG, hexa-histidine tag and a tobacco etch virus (TEV) protease cleavage site had been cloned between XbaI and BamHI restriction sites immediately upstream of the protein coding region (pET28HisTEV). An NheI site was created by PCR between the stop codon and NotI site. A truncated *PB1* gene encoding residues 1–81 was cloned into pET21b using NdeI and NotI. A co-expression plasmid was produced by ligating the PB1 coding fragment into NheI–NotI digested pET28HisTEV-PA. This co-expression plasmid was transformed into BL21(DE3)RILP CodonPlus strain (Stratagene). The protein complex was expressed in LB medium overnight at 15 °C after induction with 0.5 mM IPTG. Harvested cells were resuspended in Ni-NTA binding buffer (20 mM Tris pH 8.0, 500 mM NaCl, 500 mM urea, 25 mM imidazole and 10 mM β-mercaptoethanol) and lysed by sonication. After centrifugation, the supernatant was loaded onto Ni-NTA agarose (Qiagen) equilibrated with the same buffer. Protein was eluted by a 25–500 mM linear gradient of imidazole. Peak fractions were incubated overnight with His-tagged TEV protease at room temperature while dialysing against Ni-NTA binding buffer. After complete cleavage the sample was loaded on Ni-NTA agarose again to remove His tag, His-tagged TEV protease and minor protein contaminants. The complex was then dialysed against Q-buffer (20 mM Tris pH 8.0, 100 mM NaCl and 5 mM dithiothreitol (DTT)) and was loaded onto Q-sepharose (GE-Healthcare). The peak fraction was eluted by a 100–1,000 mM linear gradient of NaCl. Finally, the protein complex was dialysed against crystallization buffer (20 mM Tris pH 8.0 and 100 mM NaCl) and then concentrated to 15 mg ml^{-1} for crystallization using a Centriprep YM-30 concentrator (Millipore) with an exclusion size of 30 kDa. Site-specific mutagenesis was carried out by PCR. Mutant proteins were purified by the same method.

Pull-down assay. A synthetic gene encoding PB1 residues 1–14 was cloned into another modified pET28b vector, pETGSTTEV, in which GST had been inserted instead of the His-tag found in pETHisTEV. After expression by the same method used for the PA–PB1 complex mentioned above, GST-fused PB1 was purified using glutathione sepharose (GE-Healthcare) equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl and 5 mM DTT by a 25–500 mM linear gradient of reduced glutathione. For PA, cloning and purification were carried out by the same method as for the PA–PB1 complex.

A total of 2 nmol GST-tagged PB1 was incubated with 4 nmol untagged PA at room temperature overnight in reconstitution buffer containing 20 mM Tris/HCl pH 8.0, 150 mM NaCl and 5 mM DTT. A total of 20 µl glutathione sepharose resin was used for pull-down proteins. After washing by repeating centrifugation and addition of same buffer, the complex was eluted by 50 µl elution buffer (20 mM reduced glutathione in the above buffer). Proteins were analysed by SDS–polyacrylamide gel electrophoresis (5–20% gradient) and staining with Coomassie blue.

Reconstitution of model viral RNP in transfected cells. 293T cells were transfected with viral protein expression plasmids encoding PA (either wild type or mutant), PB1, PB2, NP and pHH21-vNS-Luc reporter plasmid. This plasmid carries the luciferase gene in reverse orientation sandwiched between 23-nucleotide-long 5′- and 26-nucleotide-long 3′-terminal promoter sequences of the influenza virus segment 8, placed under the control of human Pol I promoter. This system can reconstitute active viral RNP[19]. After incubation for 16 h, the luciferase assay[32] (Promega) and real-time RT–PCR assay were carried out. RNA purified from the cells was subjected to reverse transcription with different primers to assess the levels of cRNA, vRNA and mRNA. 5′-TATGAACATTTCGCAGCCTACCGTAGTGTT-3′, corresponding to the luciferase coding region between nucleotide sequence positions 351–380, was used to measure the cRNA levels. 5′-AGTAGAAACAAGGGTGTTTTTTAGTA-3′, which is complementary to the 3′ portion of the segment 8 cRNA, was used to measure vRNA synthesis, and oligo(dT)$_{20}$ was used to measure mRNA. The single-stranded cDNAs produced were subjected to real-time quantitative PCR analyses with two specific primers: 5′-TATGAACATTTCGCAGCCTACCGTAGTGTT-3′, corresponding to the luciferase coding region between nucleotide sequence positions 351–380, and 5′-CCGGAATGATTTGATTGCCA-3′, complementary to the luciferase coding region between nucleotide sequence positions 681–700. *NP* mRNA transcribed from the expression plasmid was used as an internal control.

Crystallization. Crystals were grown at 20 °C in hanging drops, which consisted of protein solution (1 µl) and reservoir solution (1 µl). For the PA–PB1 complex, 10 mg ml^{-1} protein in 20 mM Tris-HCl (pH 8.0), 100 mM sodium chloride, was crystallized with 100 mM Tris-HCl (pH 7.5), 2.4 M sodium formate reservoir solution. For heavy atom derivatization, crystals were soaked in a solution containing crystallization buffer and 0.5 mM thimerosal (Hg) for 12 h before data collection. Crystals were of space group $P3_221$, with $a = b = 101.957$ Å, $c = 115.023$ Å, and contained one molecule in the asymmetric unit. Diffraction data were collected at −180 °C using a crystal flash-frozen in crystallization buffer containing 30% (v/v) glycerol. Diffraction data from native and mercury-derivatized crystals were collected at 1.0 Å and 1.008 Å, respectively, on beamlines SPring8 BL41XU and Photon Factory BL5A, using an ADSC Quantum 315 CCD detector. Diffraction data were integrated and scaled with HKL2000 and SCALEPACK[33]. General handling of the scaled data was carried out with programs from the CCP4 suite[34].

Structure determination and refinement. The native and derivative data sets were used for phasing by SIR using SHELXD[35] and SOLVE[36]. Five mercury sites and initial phases were determined using SOLVE. Solvent flattening using RESOLVE[37] was used to improved phase accuracy. After density modification, an electron density map was calculated to a resolution of 3.2 Å. The map was of good quality, allowing most of the model to be traced readily. Successive rounds of model building using COOT[38] and TURBO-FRODO[39] were followed by refinement using CNS[40] and REFMAC[41]. Solvent molecules were placed at positions where spherical electron density peaks were found above 1.3σ in the $|2F_o - F_c|$ map and above 3.0σ in the $|F_o - F_c|$ map, and where stereochemically reasonable hydrogen bonds were allowed. The final model contained residues 257–348, 354–371, 398–549, 558–716 of PA, and 1–15 of PB1 protein. Validation of the final model was carried out using PROCHECK[42], which indicated that 87.7% of the residues are in the most favourable regions of the Ramachandran plot, with no residues in 'disallowed' regions. A summary of the data collection and refinement statistics is given in Supplementary Table 1.

31. Kawaguchi, A. & Nagata, K. *De novo* replication of the influenza virus RNA genome is regulated by DNA replicative helicase, MCM. *EMBO J.* **26**, 4566–4575 (2007).
32. Turan, K. *et al.* Nuclear MxA proteins form a complex with influenza virus NP and inhibit the transcription of the engineered influenza virus genome. *Nucleic Acids Res.* **32**, 643–652 (2004).
33. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
34. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
35. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
36. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
37. Terwilliger, T. C. SOLVE and RESOLVE: Automated structure solution and density modification. *Methods Enzymol.* **374**, 22–37 (2003).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
39. Roussel, A. & Cambillau, C. in *Silicon Graphics Geometry Partner Directory* 77–78 (Silicon Graphics, 1989).
40. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
41. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
42. Laskowski, R. A., Moss, D. S. & Thornton, J. M. Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067 (1993).

 Vol 454 | 28 August 2008 | doi:10.1038/nature07173

LETTERS

Multipotent somatic stem cells contribute to the stem cell niche in the *Drosophila* testis

Justin Voog[1,2], Cecilia D'Alterio[1] & D. Leanne Jones[1]

Adult stem cells reside in specialized microenvironments, or niches, that have an important role in regulating stem cell behaviour[1]. Therefore, tight control of niche number, size and function is necessary to ensure the proper balance between stem cells and progenitor cells available for tissue homeostasis and wound repair. The stem cell niche in the *Drosophila* male gonad is located at the tip of the testis where germline and somatic stem cells surround the apical hub, a cluster of approximately 10–15 somatic cells that is required for stem cell self-renewal and maintenance[2–4]. Here we show that somatic stem cells in the *Drosophila* testis contribute to both the apical hub and the somatic cyst cell lineage. The *Drosophila* orthologue of epithelial cadherin (DE-cadherin) is required for somatic stem cell maintenance and, consequently, the apical hub. Furthermore, our data indicate that the transcriptional repressor *escargot* regulates the ability of somatic cells to assume and/or maintain hub cell identity. These data highlight the dynamic relationship between stem cells and the niche and provide insight into genetic programmes that regulate niche size and function to support normal tissue homeostasis and organ regeneration throughout life.

Many stem cell niches include support cells that influence stem cell behaviour through secretion of diffusible molecules. Physical contact between stem cells and support cells and/or the extracellular matrix holds stem cells within the niche and close to self-renewal signals. Furthermore, niches provide spatial and mechanical cues that influence the fate of stem cell daughters. Therefore, the stem cell niche has an important role in regulating stem cell maintenance, self-renewal and survival (reviewed in ref. 5). However, little is known about the factors that regulate niche maintenance or size.

Approximately ten somatic cells, called the hub, are found at the apical tip of the *Drosophila* testis (Fig. 1a)[2]. Germline stem cells (GSCs) and somatic stem cells (SSCs) surround and are in contact with hub cells. Whereas GSCs sustain spermatogenesis, SSCs produce cyst cells that encapsulate the maturing germ cells and ensure differentiation[6,7]. Hub cells secrete the growth factor Unpaired (Upd)[3,4], which activates the JAK-STAT signal transduction pathway in adjacent stem cells. JAK-STAT signalling is necessary for stem cell maintenance and is sufficient to specify self-renewal of both GSCs and SSCs in the testis[3,4,8].

The apical hub is typically described as a post-mitotic, static structure. However, in agametic flies, SSCs proliferate and express hub markers, leading to an apparent expansion of the apical hub[9]. Furthermore, recent studies have noted that hub cell number and function decrease with age[10,11], indicating that the stem cell niche in the testis is dynamic. Hub cells and SSCs share numerous features, including similar gene expression patterns and close association with GSCs and each other[2,12]; however, the precise relationship between SSCs and hub cells has not been explored.

We proposed that SSCs may serve as a source of cells that contribute to the apical hub and, consequently, the stem cell niche. To address whether SSCs give rise to hub cells, positively marked β-galactosidase-expressing (β-gal^{+}) SSCs were generated using mitotic recombination, a technique typically used for lineage tracing

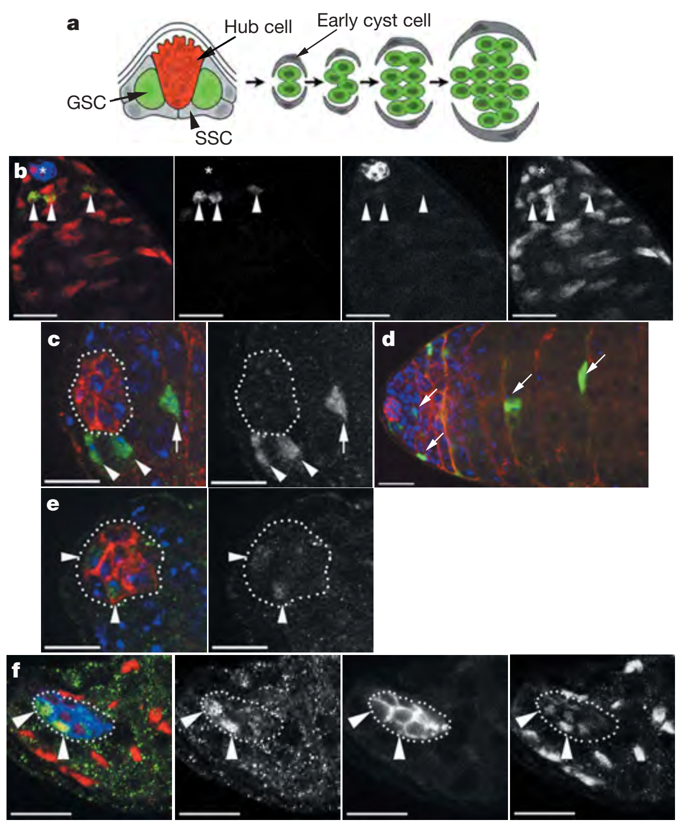


Figure 1 | Somatic stem cell progeny contribute to the hub. a, Schematic diagram of the apical tip of the testis. Hub cells (red) contact SSCs (light grey) and GSCs (green). SSCs produce cyst cells (dark grey) that envelop developing germline cells (green). **b–f**, Immunofluorescence images of the apical tip of the adult testis from flies heat shocked to generate labelled dividing cells expressing nuclear β-galactosidase (β-gal). Testes were stained with antibodies to DE-cadherin or Fasciclin III (FasIII), β-gal, and Traffic Jam (TJ) or 4,6-diamidino-2-phenylindole (DAPI). **b**, One day after heat shock, three β-gal^{+} (arrowheads, green; second panel) and TJ^{+} somatic cells (red; fourth panel) are near the FasIII^{+} (blue; third panel) hub (asterisk). **c**, Five days after heat shock two β-gal^{+} SSCs (arrowheads, green; right panel) contact hub cells (outline/DE-cadherin, red). A β-gal^{+} SSC daughter cell (arrow) is also present. **d**, Lower magnification view of **c**. β-Gal^{+} somatic cells (arrows, green) are present along the length of the testis. **e**, Five days after heat shock, the hub (outline, DE-cadherin, red) contains two β-gal^{+} cells (arrowheads, green; right panel). **f**, Fifteen days after heat shock, two β-gal^{+} hub cells (green, arrowheads; second panel) co-stain with FasIII (blue, outline; third panel) and TJ (red; fourth panel). Scale bars: **b**, **d**, 20 μm; **c**, **e**, **f**, 10 μm.

[1]Laboratory of Genetics, The Salk Institute for Biological Studies, and [2]Department of Biomedical Sciences, University of California San Diego, La Jolla, California 92037, USA.

analyses. Labelled SSCs were generated by heat-shocking flies of the appropriate genotype to initiate FLP-mediated recombination, resulting in reconstitution of the α-*tubulin* promoter upstream of the *lacZ* gene[13]. Hub cells were identified by immunolabelling with antibodies to Fasciclin III (FasIII, Fig. 1b, third panel)[14] or DE-cadherin (Fig. 1c)[15], cell-surface proteins concentrated at hub cell junctions. SSCs and early cyst cells were identified by immunolabelling with antibodies to Traffic Jam (TJ), a transcription factor that is strongly expressed in early cyst cell nuclei[16] (Fig. 1b, f, first panels) and weakly expressed in hub cells (Fig. 1f, fourth panel).

Wild-type SSC clones were identified as β-gal$^+$ cells adjacent to the apical hub and surrounding germ cells (Fig. 1c and Supplementary Table 1). A series of heat shocks after eclosion (hatching) led to at least one β-gal$^+$ SSC in 57% ($n = 47$) of testes from 3-day-old males analysed 1 day after heat shock. At 5 days after heat shock, 28% ($n = 74$) of testes contained at least one β-gal$^+$ SSC, and this frequency decreased to 10% ($n = 183$), 10% ($n = 225$) and 4% ($n = 142$) at 10, 15 and 30 days after heat shock, respectively (Supplementary Table 1).

In addition to β-gal$^+$ SSCs, β-gal$^+$ hub cells were also observed that co-labelled with DE-cadherin and FasIII. In fact, β-gal$^+$ cells were found within the hub in 34% ($n = 47$), 47% ($n = 74$) and 60% ($n = 225$) of testes from males at 1, 5 and 15 days after heat shock, respectively (Fig. 1e, f and Supplementary Table 2). No β-gal$^+$ cells were observed in flies not exposed to the heat-shock protocol (Supplementary Table 2).

Previous reports concluded that hub cells are post-mitotic[2,9]; however, it is possible that hub cells undergo rare divisions to become marked during recombination. To test whether hub cells are mitotic, dividing cells in the testis were labelled with 5′-bromo-2-deoxyuridine (BrdU), which is incorporated into newly synthesized DNA during S phase. Flies were fed ('pulsed') BrdU for 30 min and aged ('chased') for up to 15 days. Subsequently, labelled testes were co-stained with antibodies to BrdU, as well as to the hub marker FasIII. No BrdU-positive (BrdU$^+$) hub cells were detected after a 1-day chase ($n = 82$), although cells adjacent to the hub were clearly labelled (Fig. 2a).

However, BrdU$^+$ hub cells were observed 3–10 days after labelling (Fig. 2b, d). BrdU$^+$ hub cells were present in 4% ($n = 143$), 8% ($n = 103$) and 3% ($n = 96$) of testes assayed at 5, 7–8 and 10 days after labelling (Supplementary Table 3). Moreover, these BrdU$^+$ cells co-expressed FasIII (Fig. 2b) and an *upd* reporter, indicating that these cells function as hub cells (Fig. 2d). These data are consistent with previous results indicating that hub cells are post-mitotic[2,9] and support the hypothesis that mitotically active SSCs act as a source of cells that can contribute to the apical hub.

SSCs are reported to be the only dividing somatic cells in the testis; however, we observed two distinct populations of somatic cells dividing near the testis tip. One group, which constituted 69% (20 out of 29) of phospho-histone-H3-positive somatic cells ($n = 453$ testes), appeared to be immediately adjacent to the hub, similar to GSCs (Fig. 3a, b). We also observed somatic cells that were dividing 1–2 cell distances away from the hub (9 out of 29) (Fig. 3c, d). Several scenarios could explain these observations: there are two SSC populations, one which gives rise to hub cells and another that sustains the cyst cell lineage, or there are SSCs that produce both hub cells and a transient amplifying somatic cell population.

To distinguish between these possibilities, we adjusted the heat-shock regime to label, on average, only one β-gal$^+$ somatic cell and analysed the clones derived from these marked cells (Supplementary Fig. 1). Thirteen per cent of testes examined contained marked somatic cells adjacent to the hub at 1 day after heat shock ($n = 185$), which decreased to 9.5% ($n = 190$), 3.3% ($n = 212$) and 2.8% ($n = 72$) of testes at 5, 10 and 15 days after heat shock, indicating a half-life for SSCs between 5–10 days. Single, marked somatic cells displaced away from the hub were found initially in 11.3% of testes examined at 1 day after heat shock, but this number decreased to 0.5% by 5 days after heat shock, which is consistent with these cells being transient non-stem-cell clones (Fig. 3e, f and Supplementary Fig. 1). In contrast, the number of testes that contained marked hub cells increased from 11.9% at 1 day after heat shock to 25.8%, 24.5% and 18.1% at 5, 10 and 15 days after heat shock, respectively. Clones containing all three cell types were observed in 22% ($n = 18$) and 14% ($n = 7$) of testes that contained marked SSCs at 5 and 10 days after heat shock, respectively (Fig. 3g–i). From these data we conclude that multipotent SSCs self-renew and contribute to both hub and cyst cell lineages, whereas dividing cyst cells, which we call cyst progenitor cells (CPCs), expand the pool of cells capable of encapsulating newly divided gonialblasts and maturing spermatogonia to ensure terminal differentiation[6,7,17] (Supplementary Fig. 1).

To identify factors required for incorporation of cells into the apical hub, SSCs were generated that were mutant for genes expressed in both cell types: DE-cadherin, which is encoded by the *shotgun* (*shg*) gene, and the transcriptional repressor Escargot (Esg). DE-cadherin

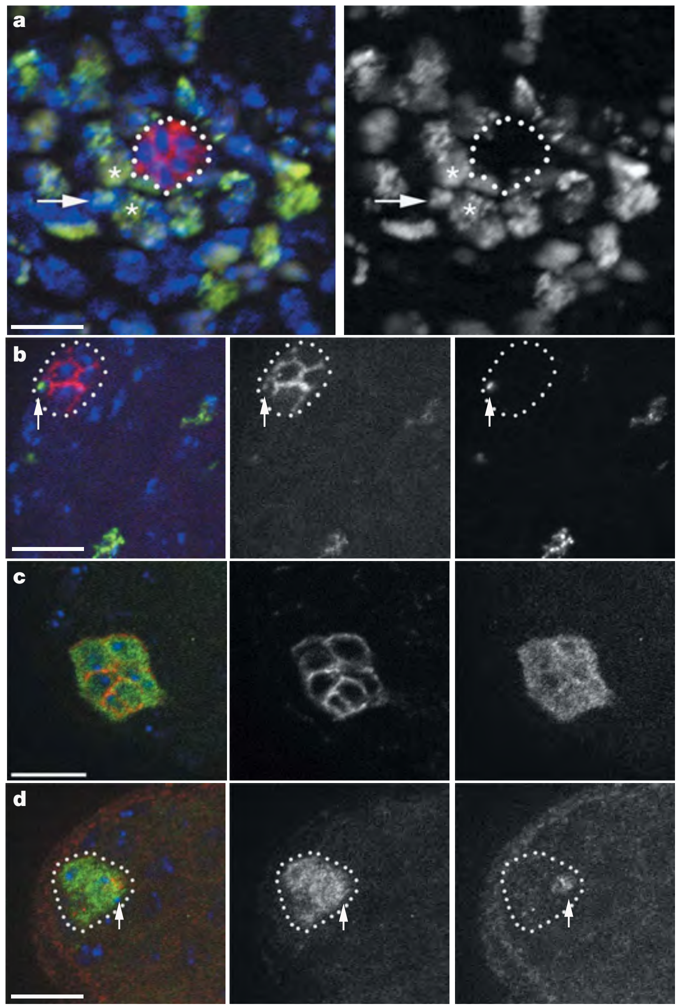


Figure 2 | BrdU-labelled cells become incorporated into the apical hub. **a**, **b**, **d**, Immunofluorescence images of testes from wild-type (*OregonR*) (**a**, **b**) or *upd*GAL4-UAS-*gfp* (**d**) adult flies fed BrdU. **a**, Left panel: BrdU$^+$ cells (green) surround the FasIII$^+$ hub (red, outline). Two BrdU$^+$ GSCs (asterisks) and a BrdU$^+$ somatic cell (arrow) are adjacent to the apical hub (outline) after a 1-day chase. Right panel: no BrdU$^+$ hub cells are detected. **b**, A BrdU$^+$ hub cell (arrow) is present after a 3-day chase (left panel). FasIII (middle panel); BrdU (right panel). **c**, Hub cells stained with DE-cadherin (red; middle panel) and GFP (green; right panel) in *upd*GAL4-UAS-*gfp* testis. **d**, A BrdU$^+$ hub cell (arrow) is detected after a 7-day chase. The middle panel shows the GFP channel; the right panel shows the BrdU channel. Scale bars: 10 μm.

is expressed in cyst cells and is strongly enriched in the hub (Fig. 1c, d). SSC clones were generated that were homozygous mutant for either the loss-of-function shg^{IG29} or amorphic shg^{IH} allele. SSC maintenance and frequency of marked hub cells were assayed at various time points. In this experiment, marked cells and their progeny subsequently become permanently labelled by ubiquitous green fluorescent protein (GFP) expression (Fig. 4).

Heat-shocked wild-type testes possessed GFP$^+$ GSCs and SSCs, as well as GFP$^+$ hub cells that co-stained with FasIII and DE-cadherin (Fig. 4 a–c, e). In contrast to wild type, *shg* mutant GSC and SSC clones were not maintained (Fig. 4d, Supplementary Fig. 2 and Supplementary Table 1), indicating that DE-cadherin has a role in stem cell maintenance in the testis, similar to its role in the ovary[18,19], presumably by holding stem cells within the niche and close to self-renewal signals.

Marked hub cells were observed in 14% ($n = 144$), 35% ($n = 114$) and 65% ($n = 110$) of wild-type testes examined at 5, 10 and 15 days after heat shock, respectively (Fig. 4e and Supplementary Table 2). Notably, progeny of DE-cadherin mutant SSCs contributed to the apical hub at a frequency similar to progeny from wild-type SSCs (Fig. 4f and Supplementary Table 2). These data indicate that although DE-cadherin is required for SSC maintenance, it is not absolutely required for mediating the contribution of SSC progeny to the hub.

To confirm that *shg* is not required in hub cells for maintaining the apical hub, we used RNAi-mediated knockdown of *shg* expression in hub cells. A FasIII$^+$ apical hub was detected in 100% of testes from 1-day-old ($n = 12$), 10-day-old ($n = 15$) and 20-day-old males ($n = 28$), despite a reduction in DE-cadherin expression in hub cells (Supplementary Fig. 3). Testes collected at 20 days also displayed normal expression of a *upd* reporter (98%, $n = 115$) and contained TJ$^+$ (100%, $n = 25$) cells near the apical tip (Supplementary Fig. 3). These data support our findings that DE-cadherin is not absolutely required in hub cells to maintain a functional stem cell niche.

However, *shg* is required in SSCs and early cyst cells for maintaining the apical hub. Knockdown of *shg* in all SSCs and early cyst cells resulted in a decrease in the number of TJ$^+$ cells in 1-day-old males, consistent with a role for *shg* in SSC maintenance (Fig. 4g, h and Supplementary Fig. 4). Surprisingly, decreased levels of DE-cadherin were also observed in hub cells (Supplementary Fig. 4). In 29% ($n = 42$) of 15-day-old and 44% ($n = 36$) of 20-day-old males, the apical hub was severely diminished or lost, as determined by FasIII expression (Fig. 4i, j and Supplementary Fig. 4). These data support our model that SSCs act as a source of cells to maintain the apical hub.

The transcriptional repressor Escargot is expressed in many tissues, including GSCs, early cyst cells and hub cells in the testis[20,21]. Males carrying a viable, hypomorphic allele of *esg*, called *shutoff*, exhibit loss of apical hub cells during development (J.V., manuscript in preparation). Therefore, we hypothesized that *esg* may be required for regulating the contribution of SSCs to the hub. Mutant labelled SSCs were generated using two amorphic *esg* alleles, as described above.

Unlike progeny from *shg* mutant SSCs, progeny from *esg* mutant SSCs did not contribute to the hub at the same frequency as wild-type controls: esg^{L2} mutant GFP$^+$ hub cells were observed in 5% ($n = 38$), 3% ($n = 40$), 0% ($n = 36$) and 4% ($n = 52$) of testes examined at 1, 5, 10 and 15 days after heat shock, respectively (Supplementary Table 2). In instances when *esg* mutant GFP$^+$ hub cells were observed, normal hub morphology was often severely disrupted (Fig. 4k, l). These data suggest that *esg* regulates either the contribution of SSC progeny to the hub, perhaps by facilitating the cell fate transition between SSC and hub cell, or maintenance of hub cell fate.

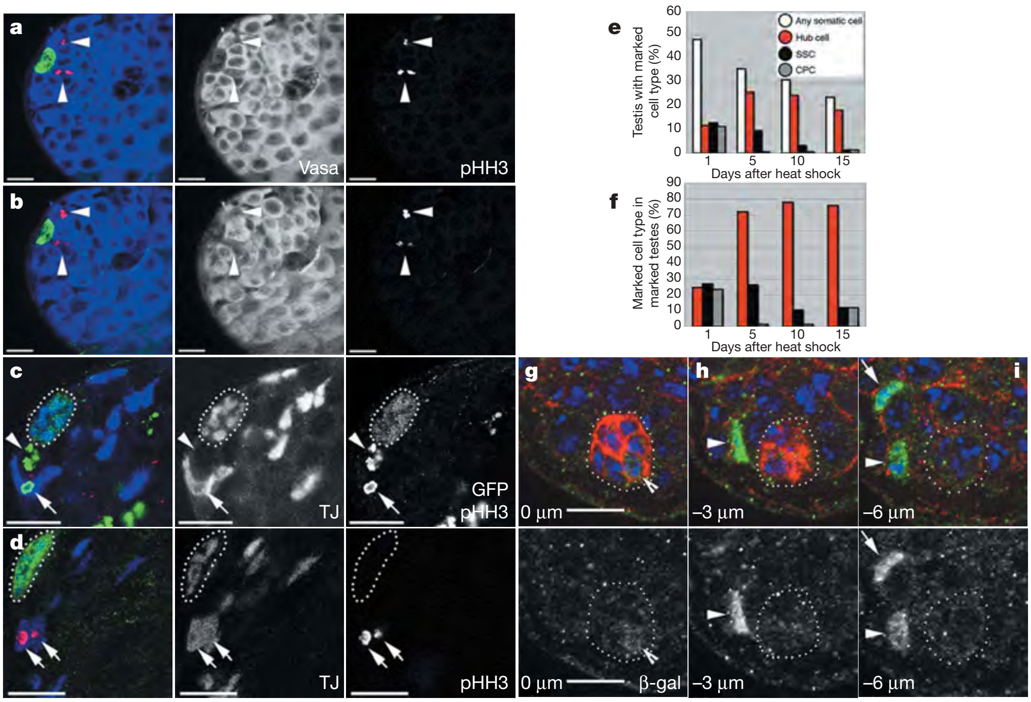


Figure 3 | Two populations of mitotically active somatic cells are present near the hub. a–d, Confocal images of *upd*GAL4-UAS-*gfp* testes. **a**, **b**, Testes stained for Vasa (blue; middle panels), phospho-histone-H3 (pHH3, red; right panels) and GFP (green). Phospho-histone-H3$^+$ SSCs (arrowheads) are observed near the hub. **c**, Testes stained for TJ (blue; middle panel), phospho-histone-H3 (green; right panel) and GFP (green; right panel). The hub is outlined. A mitotic CPC (arrow) surrounds a mitotic GSC (arrowhead). **d**, Testes stained with TJ (blue; middle panel), phospho-histone-H3 (red; right panel) and GFP (green). A mitotic CPC (arrows) one-cell distance from the hub (outline) is shown. **e**, Graph depicting marked somatic cell type frequency from all testes assayed at various time points. **f**, Graph depicting marked somatic cell type frequency from testes containing marked somatic cells at various time points. **g–i**, A testis 10 days after heat shock stained for β-gal$^+$ (green; bottom panels), DE-cadherin (red) and DAPI (blue). A β-gal$^+$ hub cell (**g**, indented arrowhead), β-gal$^+$ SSC (**h**, **i**, arrowhead) and β-gal$^+$ CPC (**i**, arrow) are shown. Z-section images (1 µm) are denoted by depth in the lower left corner. Scale bars: 10 µm.

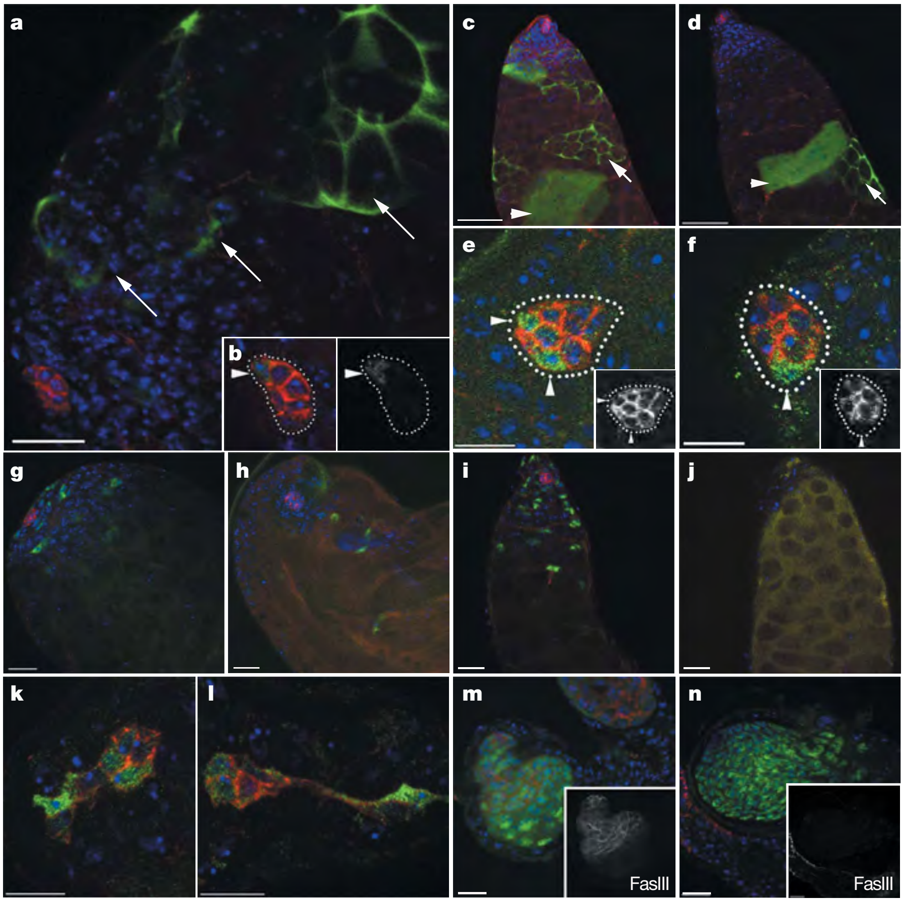


Figure 4 | Factors required for SSC maintenance and the SSC–hub cell transition. **a**, Ten days after heat shock a series of GFP$^+$ late cyst cells (green, arrows) and a single labelled hub cell are present (arrowhead in **b**). GFP (green), DE-cadherin (red) and DAPI (blue) are shown. **b**, Higher magnification (×63) of the hub in **a** (outline, DE-cadherin, red) with a GFP$^+$ hub cell (arrowhead, right panel). **c**, Five days after heat shock GFP$^+$ germline and somatic cells are maintained near the DE-cadherin$^+$ hub (red). Late GFP$^+$ germline (arrowhead) and somatic cell (arrow) clones are present basally. **d**, Five days after heat shock GFP$^+$ *shg^{IG29}* mutant late somatic cells are present (arrow), but no GFP$^+$ clones are present near the DE-cadherin$^+$ hub (red) at the apical tip. **e**, **f**, Fifteen days after heat shock GFP$^+$ hub cells are present from wild-type (**e**) and *shg^{IG29}* (**f**) somatic clones (arrowheads). GFP$^+$ *shg^{IG29}* hub cells display loss of DE-cadherin (red) at the hub interface (inset, DE-cadherin channel). **g**, **h**, Testes from 1-day-old *c587*GAL4 (**g**) and *c587*GAL4/UAS-*shgRNAi* (**h**) males stained with DE-cadherin (red) and TJ (green). **i**, **j**, Testes from 15-day-old *c587*GAL4 (**i**) and *c587*GAL4/UAS-*shgRNAi* (**j**) males. Note the loss of DE-cadherin (red) and TJ (green) stain in **j**. **k**, **l**, Fifteen days after heat shock distorted GFP$^+$ hub cells are present from *esg^{G66}* (**k**) and *esg^{L2}* (**l**) mutant somatic clones (compare to **e**). The hub is marked with DE-cadherin (red). **m**, **n**, *esg^{shof}/CyO; osk/osk* (**m**) and *esg^{shof}/esg^{shof}; osk/osk* (**n**) testes stained for FasIII (red, insets) and TJ (green). Scale bars: **a**, **c**, **d**, 20 μm; **e**–**n**, 10 μm.

To explore Esg function further, we used the agametic *oskar* (*osk*) mutant phenotype, in which SSCs proliferate and express hub markers, resulting in an apparent expansion of the apical hub[9]. If Esg is required for mediating the transition of somatic cyst cells to the apical hub, we predicted that the expansion of FasIII$^+$ cells would be blocked in an *esg*;*osk* double mutant background. In contrast to the expansion of FasIII$^+$ cells in 82% ($n = 16$) of *osk* mutant testes, only 22% ($n = 63$) of testes from *esg^{shof}*;*osk* mutant males showed expansion of FasIII, despite there being clearly more TJ$^+$ somatic cyst cells (Fig. 4m, n). These data support our previous results and indicate that *esg* is required for the ability of somatic cells to assume and/or maintain hub cell fate.

Our findings demonstrating that SSCs can adopt a hub cell fate highlight the dynamic nature of the stem cell–niche relationship and provide a mechanism to regulate the size and function of the stem cell niche in the *Drosophila* testis. In our model, as somatic cells are displaced from the hub, there is a decline in self-renewal and proliferation potential, which could be reinforced by encapsulation of differentiating germ cells (Supplementary Fig. 1). Interestingly, expansion of the somatic cyst cells as a consequence of germline loss suggests that germ cells exert an anti-proliferative influence that must be overcome in SSCs.

A better understanding of how stem cell niches are established and regulated in mammalian systems could facilitate modulation of the niche to enhance transplantation of stem cells in regenerative medicine[22,23]. Conversely, if an expanded or modified niche accompanies tumour progression or metastasis, then blocking niche maintenance programmes (niche ablation) could be used as an important anti-cancer therapeutic[24,25].

METHODS SUMMARY

Fly husbandry and stocks. Flies were raised at 25 °C on standard cornmeal-molasses media. The *esg^{shof}* allele was isolated in the Fuller laboratory (J.V. *et al.*, manuscript in preparation). The UAS-*shgRNAi* line was obtained from the Vienna *Drosophila* Resource Center (VDRC). The *upd*GAL4-UAS-*gfp* line was a gift from E. Bach. For additional information see Methods.

Lineage tracing analyses. Adult flies or embryos (*y,w, hs-FLP/+*; *X-15-29/X-15-33*) were heat shocked for 2 h on 2 consecutive days (maximal clone induction) or for 1 h on 1 day (moderate clone induction)[26].

For MARCM experiments, late pupae or early larvae were heat shocked at 37 °C for 2 h on 2 consecutive days. Flies were collected at various time points after the last heat shock, immunostained and analysed[27].

BrdU labelling. *OregonR* or *upd*GAL4-UAS-*gfp* 1-day-old males were starved in wet-plugged vials overnight (~16 h). Subsequently, starved flies were moved to vials containing 100 μl of 100 mM BrdU dissolved in grape juice for 30 min and

then moved to agar-molasses vials and aged. Testes were stained as described previously[11].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 February; accepted 16 June 2008.
Published online 20 July 2008.

1. Schofield, R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* **4,** 7–25 (1978).
2. Hardy, R. W., Tokuyasu, K. T., Lindsley, D. L. & Garavito, M. The germinal proliferation center in the testis of *Drosophila melanogaster. J. Ultrastruct. Res.* **69,** 180–190 (1979).
3. Kiger, A. A., Jones, D. L., Schulz, C., Rogers, M. B. & Fuller, M. T. Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue. *Science* **294,** 2542–2545 (2001).
4. Tulina, N. & Matunis, E. Control of stem cell self-renewal in *Drosophila* spermatogenesis by JAK-STAT signaling. *Science* **294,** 2546–2549 (2001).
5. Jones, D. L. & Wagers, A. J. No place like home: anatomy and function of the stem cell niche. *Nature Rev. Mol. Cell. Biol.* **9,** 11–21 (2008).
6. Kiger, A. A., White-Cooper, H. & Fuller, M. T. Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature* **407,** 750–754 (2000).
7. Tran, J., Brenner, T. J. & DiNardo, S. Somatic control over the germline stem cell lineage during *Drosophila* spermatogenesis. *Nature* **407,** 754–757 (2000).
8. Harrison, D. A., McCoon, P. E., Binari, R., Gilman, M. & Perrimon, N. *Drosophila* unpaired encodes a secreted protein that activates the JAK signaling pathway. *Genes Dev.* **12,** 3252–3263 (1998).
9. Gönczy, P. & DiNardo, S. The germ line regulates somatic cyst cell proliferation and fate during *Drosophila* spermatogenesis. *Development* **122,** 2437–2447 (1996).
10. Wallenfang, M. R., Nayak, R. & DiNardo, S. Dynamics of the male germline stem cell population during aging of *Drosophila melanogaster. Aging Cell* **5,** 297–304 (2006).
11. Boyle, M., Wong, C., Rocha, M. & Jones, D. L. Decline in self-renewal factors contributes to aging of the stem cell niche in the *Drosophila* testis. *Cell Stem Cell* **1,** 470–478 (2007).
12. Boyle, M. & DiNardo, S. Specification, migration and assembly of the somatic cells of the *Drosophila* gonad. *Development* **121,** 1815–1825 (1995).
13. Harrison, D. A. & Perrimon, N. Simple and efficient generation of marked clones in *Drosophila. Curr. Biol.* **3,** 424–433 (1993).
14. Gönczy, P., Viswanathan, S. & DiNardo, S. Probing spermatogenesis in *Drosophila* with P-element enhancer detectors. *Development* **114,** 89–98 (1992).
15. Yamashita, Y., Jones, D. L. & Fuller, M. T. Orientation of asymmetric stem cell division by the APC tumor suppressor and centrosome. *Science* **301,** 1547–1550 (2003).
16. Li, M. A., Alls, J. D., Avancini, R. M., Koo, K. & Godt, D. The large Maf factor Traffic Jam controls gonad morphogenesis in *Drosophila. Nature Cell Biol.* **5,** 994–1000 (2003).
17. Schulz, C., Wood, C. G., Jones, D. L., Tazuke, S. I. & Fuller, M. T. Signaling from germ cells mediated by the rhomboid homolog stet organizes encapsulation by somatic support cells. *Development* **129,** 4523–4534 (2002).
18. Song, X., Zhu, C. H., Doan, C. & Xie, T. Germline stem cells anchored by adherens junctions in the *Drosophila* ovary niches. *Science* **296,** 1855–1857 (2002).
19. Song, X. & Xie, T. DE-cadherin-mediated cell adhesion is essential for maintaining somatic stem cells in the *Drosophila* ovary. *Proc. Natl Acad. Sci. USA* **99,** 14813–14818 (2002).
20. Streit, A., Bernasconi, L., Sergeev, P., Cruz, A. & Steinmann-Zwicky, M. mgm 1, the earliest sex-specific germline marker in *Drosophila*, reflects expression of the gene esg in male stem cells. *Int. J. Dev. Biol.* **46,** 159–166 (2002).
21. Le Bras, S. & Van Doren, M. Development of the male germline stem cell niche in *Drosophila. Dev. Biol.* **294,** 92–103 (2006).
22. Conboy, I. M. *et al.* Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature* **433,** 760–764 (2005).
23. Adams, G. B. *et al.* Therapeutic targeting of a stem cell niche. *Nature Biotechnol.* **25,** 238–243 (2007).
24. Kaplan, R. N. *et al.* VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438,** 820–827 (2005).
25. Sneddon, J. B. *et al.* Bone morphogenetic protein antagonist gremlin 1 is widely expressed by cancer-associated stromal cells and can promote tumor cell proliferation. *Proc. Natl Acad. Sci. USA* **103,** 14842–14847 (2006).
26. Nystul, T. & Spradling, A. An epithelial niche in the *Drosophila* ovary undergoes long-range stem cell replacement. *Cell Stem Cell* **1,** 277–285 (2007).
27. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22,** 451–461 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Bach, D. Godt, S. Hayahsi, N. Perrimon and R. Read for reagents and fly stocks, and Jones laboratory members, G. Adams, M. Buszczak, C. Schulz, S. DiNardo and M. Fuller for discussions and comments on the manuscript. This work was supported by a training grant from the California Institute for Regenerative Medicine to the University of California-San Diego (L. Goldstein). D.L.J. is funded by the Ellison Medical Foundation, the American Federation for Aging Research, the G. Harold and Leila Y. Mathers Charitable Foundation, the ACS and the NIH.

Author Contributions J.V. and D.L.J. planned experiments; J.V. and C.D'A. performed experiments and analysed data; D.L.J. wrote the manuscript; and J.V. and C.D'A. edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.L.J. (ljones@salk.edu).

METHODS

Fly stocks. The following stocks were used; more information can be found at Flybase (http://flybase.bio.indiana.edu): *hs-FLP-12*, *X-15-29*, *X-15-33*, *OregonR*, *hs-FLP-122*, α-*tubulin*GAL80, α-*tubulin*GAL4, *2x*-UAS*eGFP*, *FRT40A*, *FRTG13*, osk^{CE4}, osk^{301}, shg^{IG29}, shg^{IH}, esg^{G66} and esg^{L2}. Flies of the genotype esg^{shof}; *osk* were generated by crossing esg^{shof}/*CyO*; osk^{CE4}/osk^{301} females to esg^{shof}/*CyO*; osk^{CE4}/*TM6b*, *Hu* males. Controls were progeny from esg^{shof}/*CyO*; osk^{CE4}/*TM6b*, *Hu* or esg^{shof}/*CyO*; osk^{301}/*TM6b*, *Hu* females crossed to esg^{shof}/*CyO*; osk^{CE4}/*TM6b*, *Hu* males. The UAS-*shgRNAi* stock was obtained from the Vienna *Drosophila* RNAi Center.

MARCM analysis. MARCM wild-type control genotypes used were: *y,w*, $hsflp^{122}$; *FRT40A/FRT40A tub*GAL80; *tub*GAL4/*2x-UAS-eGFP* and *y,w*, $hsflp^{122}$; *FRTG13/FRTG13 tub*GAL80; *tub*GAL4/*2x-UAS-eGFP*. Mutant genotypes used were: esg^{G66} (y,*w*, $hsflp^{122}$; *FRT40A* esg^{G66}/*FRT40A tub*GAL80; *tub*GAL4/*2x-UAS-eGFP*), esg^{L2} (*y,w*, $hsflp^{122}$; *FRT40A* esg^{L2}/*FRT40A tub*GAL80; *tub*GAL4/*2x-UAS-eGFP*), shg^{IG29} (*y,w*, $hsflp^{122}$; *FRTG13* shg^{IG29}/*FRTG13 tub*GAL80; *tub*GAL4/*2x-UAS-eGFP*), shg^{IH} (*y,w*, $hsflp^{122}$; *FRTG13* shg^{IH}/*FRTG13 tub*GAL80; *tub*GAL4/*2x-UAS-eGFP*).

Immunostaining and microscopy. Testes were dissected and fixed in 2% PFA in PLP buffer (0.075 M lysine, 0.01 M sodium phosphate buffer pH 7.4) for 60 min at room temperature, rinsed for 30 min in PBS containing 0.3% Triton X-100 and 0.3% sodium deoxycholate, rinsed for 15 min in PBS containing 0.1% Triton X-100, blocked for 30 min in PBS containing 0.1% Triton X-100 and 0.3% bovine serum albumin, and immunostained with appropriate antibodies overnight at 4 °C. Testes were then washed for 60 min at room temperature in PBS containing 0.1% Triton X-100, and incubated with appropriate secondary antibodies at room temperature for 4 h. Testes were analysed with a Zeiss Axiovert 200 microscope and processed, and AxioVision (version 4.5; C. Zeiss) and Adobe Photoshop software. Analysis of FasIII-, TJ- and β-galactosidase-labelled testes were performed using a Leica TCS SP2 AOBS confocal microscope and LCS Lite 2.61.1537 software (Leica Microsystems).

Antibodies. Testes were stained using the following: rabbit anti-β-galactosidase (1:2,000) (Cappel), guinea-pig anti-TJ (1:3,000) (gift from D. Godt), mouse anti-BrdU (1:100) (BD Biosciences), rat-anti-BrdU (1:100) (Accurate Chemicals), rabbit anti-GFP (1:5,000) (Molecular Probes), and rabbit anti-phosphorylated histone H3 (1:200) (Upstate Biotechnologies). Mouse anti-Fasciclin III (7G10) (1:10) and rat anti-DE-cadherin (DCAD2) (1:20) were obtained from the Developmental Studies Hybridoma Bank. Secondary antibodies were diluted 1:500 (Molecular Probes). Samples were mounted in Vectashield mounting medium with DAPI (Vector Laboratories).

LETTERS

Mouse development with a single E2F activator

Shih-Yin Tsai[1,2,3]*, Rene Opavsky[1,2,3]*, Nidhi Sharma[1,2,3]*, Lizhao Wu[1,2,3]†, Shan Naidu[2,3,4], Eric Nolan[1,2,3], Enrique Feria-Arias[1,2,3], Cynthia Timmers[1,2,3], Jana Opavska[1,2,3], Alain de Bruin[1,2,3], Jean-Leon Chong[1,2,3], Prashant Trikha[1,2,3], Soledad A. Fernandez[5], Paul Stromberg[4], Thomas J. Rosol[4] & Gustavo Leone[1,2,3]

The E2F family is conserved from *Caenorhabditis elegans* to mammals, with some family members having transcription activation functions and others having repressor functions[1,2]. Whereas *C. elegans*[3] and *Drosophila melanogaster*[4,5] have a single E2F activator protein and repressor protein, mammals have at least three activator and five repressor proteins[1,2,6]. Why such genetic complexity evolved in mammals is not known. To begin to evaluate this genetic complexity, we targeted the inactivation of the entire subset of activators, *E2f1*, *E2f2*, *E2f3a* and *E2f3b*, singly or in combination in mice. We demonstrate that *E2f3a* is sufficient to support mouse embryonic and postnatal development. Remarkably, expression of *E2f3b* or *E2f1* from the *E2f3a* locus (*E2f3a*^{3bki} or *E2f3a*^{1ki}, respectively) suppressed all the postnatal phenotypes associated with the inactivation of *E2f3a*. We conclude that there is significant functional redundancy among activators and that the specific requirement for *E2f3a* during postnatal development is dictated by regulatory sequences governing its selective spatiotemporal expression and not by its intrinsic protein functions. These findings provide a molecular basis for the observed specificity among E2F activators during development.

Since the identification of the founding E2F family member, *E2f1* (ref. 7), two distinct genes in nematodes and flies and eight genes in mammals have been identified to encode the signature DNA binding domain that endow these transcription factors with E2F classification[1,2,6]. Among the mammalian E2F activator subset, the *E2f3* gene has emerged as the critical family member involved in the control of cell proliferation and development[8,9]. The *E2f3* locus was originally thought to encode a single DNA binding activity, but was later shown to drive the expression of two related isoforms, *E2f3a* and *E2f3b*, from two distinct promoters[10]. Given the critical link between the *E2f3* locus and the control of cell proliferation, we used homologous recombination to disrupt individually its two isoforms in mice and rigorously evaluate how their functions are integrated with that of other E2F activators. The inactivation of *E2f3a* or *E2f3b* was achieved by targeting exon 1a or 1b sequences, respectively, using Cre/*loxP* technology (Fig. 1a). Mice deleted for either exon 1a or exon 1b were identified by Southern blot and genomic polymerase chain reaction (PCR) analysis (Fig. 1b). Specific ablation of *E2f3a* or *E2f3b* was confirmed by western blot assays using total E2F3-specific antibodies (Fig. 1c).

It was previously shown that inactivation of both *E2f3a* and *E2f3b* (*E2f3*$^{-/-}$) in mice with a mixed strain background yielded offspring that developed rather normally[8,9], but we show here that breeding these mice into a pure strain background (~98% pure) resulted in embryonic lethality (Fig. 1e and Supplementary Fig. 1). Intercrossing *E2f3*$^{+/-}$ mice of different pure backgrounds restored viability of

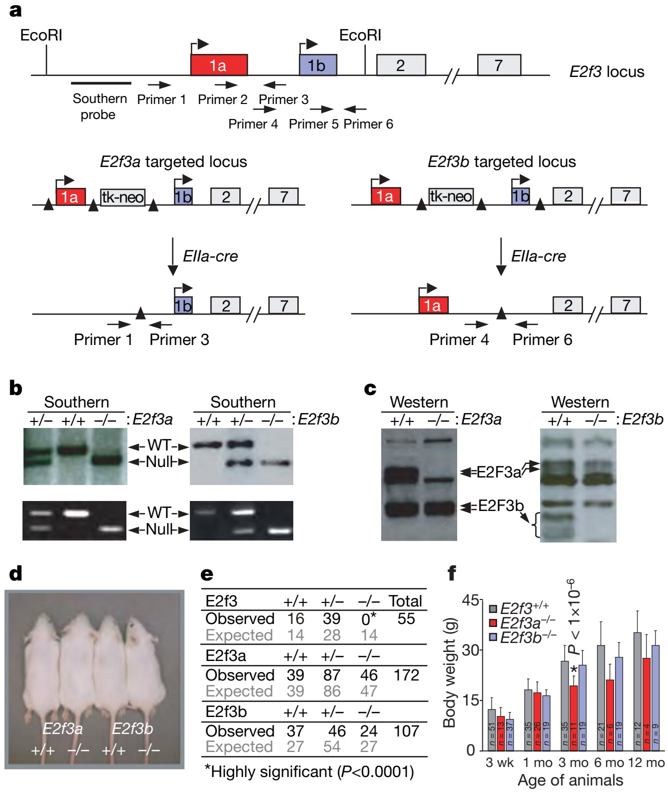


Figure 1 | Generation of *E2f3a* and *E2f3b* knockout mice. a, Partial exon/intron structure of the mouse *E2f3* gene. The schematic illustrates the *E2f3* locus with two separate promoters driving the expression of *E2f3a* and *E2f3b*. The two transcription start sites are indicated by bent arrows. The targeting vectors and the final *E2f3a* and *E2f3b* targeted alleles are shown on the left and right, respectively. The solid bar represents a HindIII-BamHI fragment used as the Southern probe. The *loxP* sites are indicated as solid triangles; PCR primers are indicated by arrows. **b**, Southern blot (top panels) and PCR genotyping assays (bottom panels) were performed on MEFs with the indicated genotypes. **c**, Western blot analysis using E2F3-specific antibodies was performed on lysates derived from MEFs with the indicated genotypes. **d**, Micrograph of 1-month-old mice with the indicated genotypes. **e**, Genotypic analysis of offspring derived from *E2f3*$^{+/-}$, *E2f3a*$^{+/-}$ and *E2f3b*$^{+/-}$ intercrosses (fifth generation FVB). A Fisher exact probability test was performed on each genetic group in comparison to the wild-type group; highly significant results are indicated by an asterisk. **f**, Body weights of *E2f3*$^{+/+}$ (grey), *E2f3a*$^{-/-}$ (red) and *E2f3b*$^{-/-}$ (blue) mice at the indicated ages; *n*, number of animals measured for each genotype at each age; error bars represent deviations within each genetic group. A Student's *t*-test was performed on each genetic group in comparison to the wild-type group. Highly significant results are indicated by an asterisk.

[1]Department of Molecular Genetics, College of Biological Sciences, [2]Human Cancer Genetics Program, Comprehensive Cancer Center, [3]Department of Molecular Virology Immunology and Medical Genetics, College of Medicine, [4]Department of Veterinary Biosciences, College of Veterinary Medicine, and [5]Center for Biostatistics, The Ohio State University, Columbus, Ohio 43210, USA. †Present address: Department of Cell Biology and Molecular Medicine and University Hospital Cancer Center, UMDNJ-New Jersey Medical School, Newark, New Jersey 07103, USA.
*These authors contributed equally to this work.

E2f3$^{-/-}$ mice, albeit with some observed strain-specific biases (Supplementary Fig. 1). In contrast, the ablation of individual *E2f3* isoforms (*E2f3a*$^{-/-}$ or *E2f3b*$^{-/-}$) in a pure strain background yielded live pups at the expected mendelian frequency (Fig. 1d–f). These results suggest that *E2f3a* and *E2f3b* have redundant functions during

a

Genotype	3a+/+ (E2f1+/+)	3a+/− (E2f1+/+)	3a−/− (E2f1+/+)	3a+/+ (E2f1+/−)	3a+/− (E2f1+/−)	3a−/− (E2f1+/−)	3a+/+ (E2f1−/−)	3a+/− (E2f1−/−)	3a−/− (E2f1−/−)	Total
Newborn	25	54	37	32	172	114	32	83	39(5)	624
Expected	29	72	43	35	156	99	35	78	49	
21 days	23	54	36	59	164	110	27	74	19*	566
Expected	25	65	39	57	142	91	32	71	45	

Genotype	3a+/+ (E2f2+/+)	3a+/− (E2f2+/+)	3a−/− (E2f2+/+)	3a+/+ (E2f2+/−)	3a+/− (E2f2+/−)	3a−/− (E2f2+/−)	3a+/+ (E2f2−/−)	3a+/− (E2f2−/−)	3a−/− (E2f2−/−)	Total
21 days	0	5	2	9	14	9	4	11	7	61
Expected	2	6	3	6	14	8	3	11	8	

Genotype	E2f3a+/+ E2f1−/−E2f2−/−	E2f3a+/− E2f1−/−E2f2−/−	E2f3a−/− E2f1−/−E2f2−/−	Total
E13.5	–	5	9	161
Expected	–	19	17	
E17.5	0	1	4	38
Expected	0	1	4	
E18.5	0	1	2(3)	43
Expected	0	1	5	
Newborn	3	11	0*(4)	275
Expected	3	13	13	

*Highly significant ($P<0.05$)

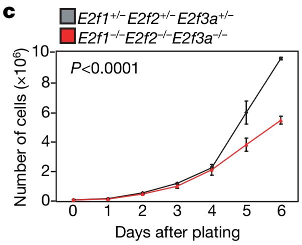
b

Genotype	3b+/+ (E2f1+/+)	3b+/− (E2f1+/+)	3b−/− (E2f1+/+)	3b+/+ (E2f1+/−)	3b+/− (E2f1+/−)	3b−/− (E2f1+/−)	3b+/+ (E2f1−/−)	3b+/− (E2f1−/−)	3b−/− (E2f1−/−)	Total
21 days	27	39	18	45	76	34	17	48	24	328
Expected	21	41	21	41	82	41	21	41	21	

Genotype	3b+/+ (E2f2+/+)	3b+/− (E2f2+/+)	3b−/− (E2f2+/+)	3b+/+ (E2f2+/−)	3b+/− (E2f2+/−)	3b−/− (E2f2+/−)	3b+/+ (E2f2−/−)	3b+/− (E2f2−/−)	3b−/− (E2f2−/−)	Total
21 days	11	17	10	20	34	16	11	24	8	156
Expected	8	18	10	18	39	21	10	21	12	

Genotype	E2f3b+/+ E2f1−/−E2f2−/−	E2f3b+/− E2f1−/−E2f2−/−	E2f3b−/− E2f1−/−E2f2−/−	Total
E13.5	–	1	2	58
Expected	–	1	3	
Newborn	0	0	2	45
Expected	0	1	2	

c



d

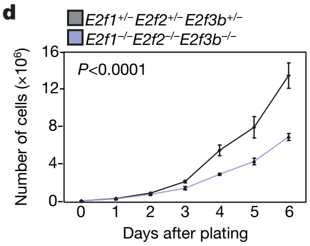


Figure 2 | Genotypic analysis of embryos and offspring deficient for various combinations of activating E2Fs. a, Genotypic analysis of offspring deficient for *E2f1*, *E2f2* and *E2f3a* (*3a*); highly significant results are indicated by an asterisk. Numbers in parentheses represent dead embryos/pups. **b**, Genotypic analysis of offspring deficient for *E2f1*, *E2f2* and *E2f3b* (*3b*). **c**, Growth curves of *E2f1*$^{-/-}$*E2f2*$^{-/-}$*E2f3a*$^{-/-}$ (red) and littermate control MEFs (grey). Four independent primary MEF lines of each genotype were plated in duplicate and counted every day for 6 days; a representative experiment is presented in **c**, where the error bars represent deviations between duplicate measurements within the same experiment. **d**, Growth curves of *E2f1*$^{-/-}$*E2f2*$^{-/-}$*E2f3b*$^{-/-}$ (blue) and littermate control MEFs (grey). Two independent primary MEF lines of each genotype were plated in duplicate and counted every day for 6 days; the entire experiment was performed twice and a representative experiment is presented in **d**, where the error bars represent deviations between duplicate measurements within the same experiment. For the cell proliferation studies in **c**, **d**, a generalized linear model (GLM) was used to study the association between cell proliferation and group behaviour, combining the data from all experiments in each genetic study. The *P*-value shown in the graphs corresponds to the main 'group effect' (see Methods).

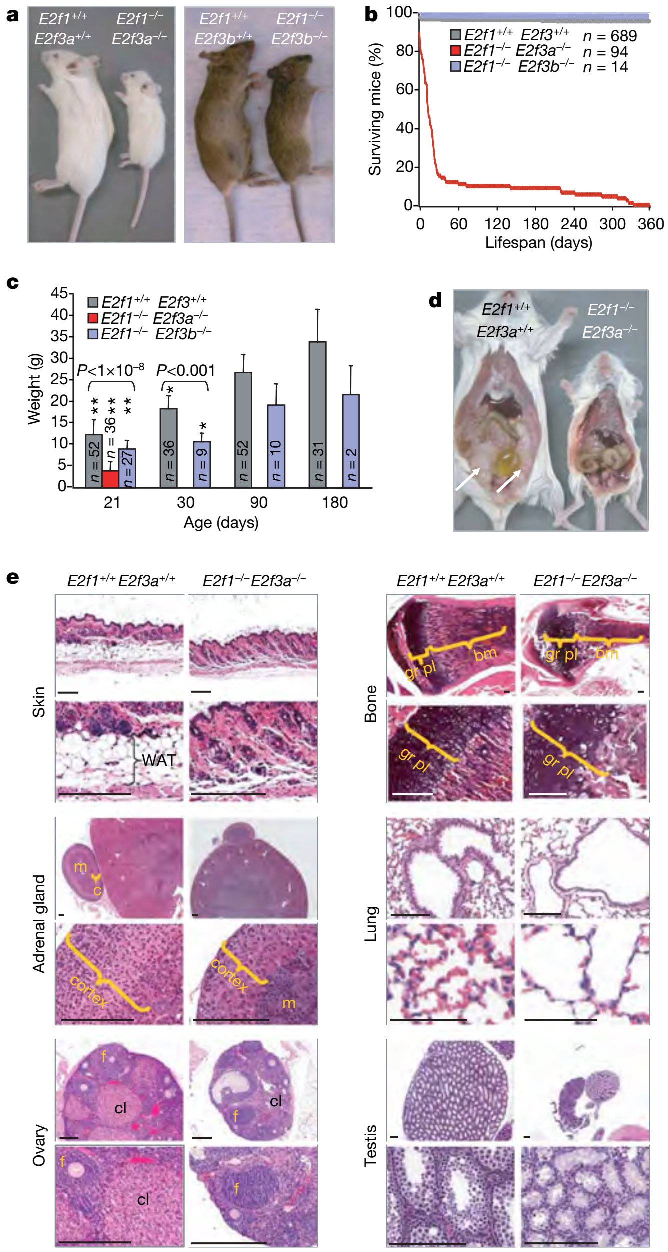


Figure 3 | *E2f1* and *E2f3a* are essential for postnatal development. a, Micrograph of *E2f1*$^{+/+}$*E2f3a*$^{+/+}$, *E2f1*$^{-/-}$*E2f3a*$^{-/-}$, *E2f1*$^{+/+}$*E2f3b*$^{+/+}$ and *E2f1*$^{-/-}$*E2f3b*$^{-/-}$ 30-day-old mice. **b**, Survival graph of mice with the indicated genotypes over a period of 1 year. **c**, Body weight of mice with the indicated genotypes and ages; *E2f1*$^{+/+}$*E2f3*$^{+/+}$ represent mice that were pooled from the *E2f1*$^{+/+}$*E2f3a*$^{+/+}$ and *E2f1*$^{+/+}$*E2f3b*$^{+/+}$ genetic groups; *n*, number of animals measured for each genotype; error bars represent deviations within each genetic group. A Student's *t*-test was performed on each genetic group in comparison to the wild-type group. Significant and highly significant results are indicated by a single and double asterisk, respectively. **d**, Micrograph of *E2f1*$^{+/+}$*E2f3a*$^{+/+}$ and *E2f1*$^{-/-}$*E2f3a*$^{-/-}$ 30-day-old mice with an open abdominal cavity; Inguinal WAT in the wild-type mouse is indicated by the white arrows; note the absence of WAT in the mutant mouse. **e**, Haematoxylin and eosin stained tissue sections of 21-day-old (unless indicated) *E2f1*$^{+/+}$*E2f3a*$^{+/+}$ and *E2f1*$^{-/-}$*E2f3a*$^{-/-}$ mice. Top left: skin sections; note the complete absence of WAT beneath the skin of the mutant mouse. Middle left: adrenal gland sections; note the reduction of the adrenal cortex in double knockout mutant mice. Bottom left: ovary sections; note the presence of follicles (f) in both genetic groups but the poor development of corpus lutea (cl) in double knockout mutant female mice (6 months of age). Top right: proximal tibia sections; note the disorganization of nuclei in the growth plate (gr pl) of double knockout mutant mice. Middle right: lung sections; note the reduction of alveolar branching in double knockout mutant mice. Bottom right: testis sections; note the reduction in testis size and lack of spermatocytes. Scale bars, 200 μm. bm, bone marrow; c, cortex; m, medulla.

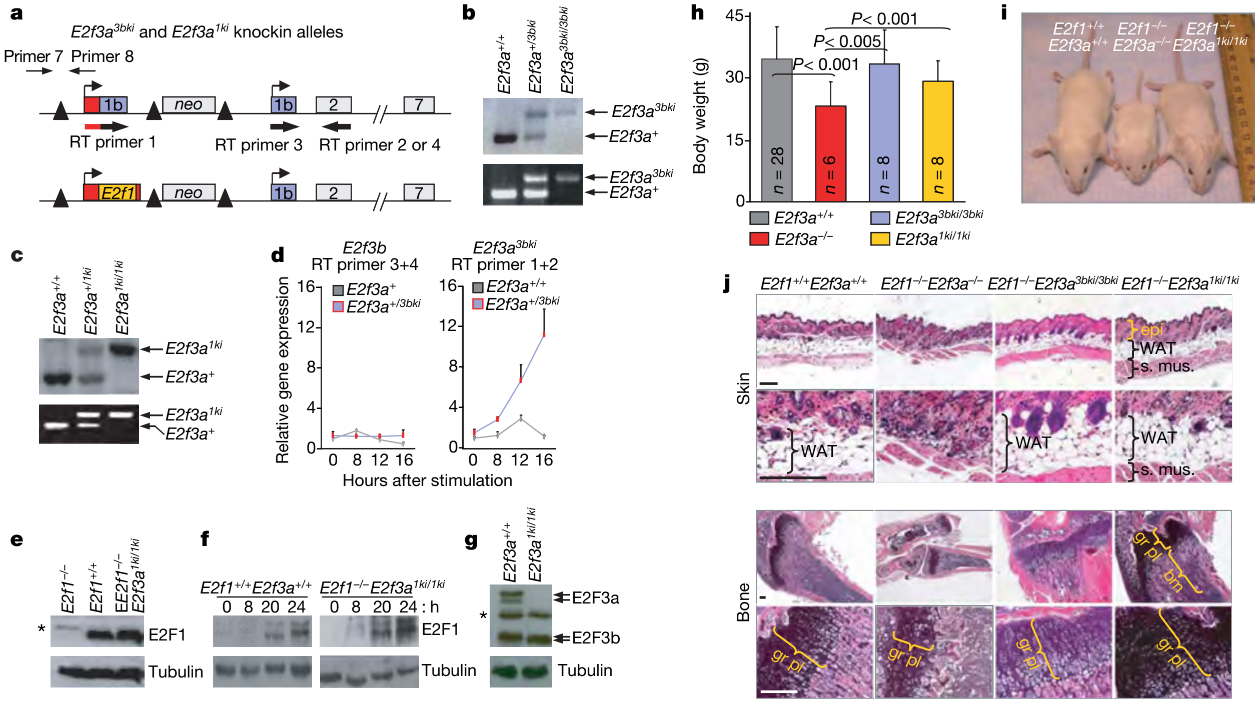


Figure 4 | Expression of *E2f3b* or *E2f1* from the *E2f3a* locus suppresses phenotypes owing to loss of *E2f3a*. **a**, Targeting strategies for replacement of *E2f3a* with *E2f3b* ($E2f3a^{3bki}$; top schematic) or *E2f3a* with *E2f1* ($E2f3a^{1ki}$; bottom schematic). Relevant exons are labelled; *E2f1* indicates the *E2f1* ORF. *loxP* sites are indicated as solid triangles; PCR genotyping primers are indicated by thin arrows; primers for measuring the expression of $E2f3a^{3bki}$ and *E2f3b* alleles by real-time (RT) PCR are indicated by thick arrows. Note that RT primer 1 spans a region that is complementary to exon 1a and exon 1b, as indicated by the red/black coloured arrow. **b**, **c**, Southern blot (top panels) and PCR (bottom panels) genotyping of genomic DNA derived from livers of 1-month-old $E2f3a^{3bki}$ (**b**) and $E2f3a^{1ki}$ (**c**) mice. **d**, Real-time PCR gene expression analysis of the *E2f3b* (left panel) and $E2f3a^{3bki}$ (right panel) alleles in $E2f3a^{+/+}$ (grey square) or $E2f3a^{+/3bki}$ (red/blue square) MEFs. Primers used for each allele are indicated (for primer sequence information see Supplementary Fig. 6). Note that RT primers used to detect $E2f3a^{3bki}$ also detect endogenous *E2f3b* allele as the 3′ portion of RT primer 1 is complementary with exon 1b sequences. Error bars are derived from reactions performed in triplicate. **e**, Western blot analysis of E2F1 protein in proliferating MEFs with the indicated genotypes; α-tubulin was detected as a loading control. Asterisk indicates a nonspecific band. **f**, Western blot analysis of E2F1 proteins in serum stimulated quiescent MEFs having the indicated genotypes; 'h' indicates the time when cells were harvested after stimulation. **g**, Western blot analysis of E2F3a and E2F3b proteins (E2F3 Sc-878) in proliferating MEFs with the indicated genotypes. Asterisk indicates a nonspecific band. **h**, Body weights of 6-month-old $E2f3a^{+/+}$, $E2f3a^{-/-}$, $E2f3a^{3bki/3bki}$ and $E2f3a^{1ki/1ki}$ mice; *n*, number of animals measured for each genotype; error bars represent deviations within each genetic group. Student's *t*-tests were performed between all genetic groups and only significant pair-wise comparisons between groups are indicated by brackets (*P*-values). **i**, Micrograph of 1-month-old $E2f3a^{+/+}$, $E2f1^{-/-}E2f3a^{-/-}$ and $E2f1^{-/-}E2f3a^{1ki/1ki}$ mice. **j**, Haematoxylin and eosin staining of tissue sections of skin and bone from 21-day-old mice with the indicated genotypes. Scale bars, 200 μm. bm, bone marrow; epi, epidermis; gr pl, growth plate; s mus, skeletal muscle.

embryonic development. Because of the functional plasticity that exists among E2F family members[8,11–13], we examined the consequences of inactivating *E2f3a* or *E2f3b* in combination with the other two known E2F activators, *E2f1* and *E2f2*. From the intercrosses described in Fig. 2a, b, we obtained pups doubly deleted for every possible combination of *E2f1*, *E2f2*, *E2f3a* or *E2f1*, *E2f2*, *E2f3b* at or near the expected ratios. Embryos with the two possible triple knockout genotypes ($E2f1^{-/-}E2f2^{-/-}E2f3a^{-/-}$ and $E2f1^{-/-}E2f2^{-/-}E2f3b^{-/-}$) were also obtained at late stages of embryonic development. Given that ablation of all four of these E2F genes in mixed strain background results in early embryonic lethality[8] (G.L., unpublished data), our current findings demonstrate that expression of at least one of the two *E2f3* isoforms is necessary and sufficient to support fetal development in the absence of other E2F activators.

The lack of any obvious proliferative defect in isoform-specific or triple knockout embryos prompted us to examine proliferation more closely in fibroblasts derived from these embryos. To this end, we generated mouse embryo fibroblasts (MEFs) from embryonic day (e)13.5 $E2f3a^{-/-}$, $E2f3b^{-/-}$, $E2f1^{-/-}E2f2^{-/-}E2f3a^{-/-}$ and $E2f1^{-/-}E2f2^{-/-}E2f3b^{-/-}$ embryos and found that all four groups of mutant MEFs had the capacity to proliferate, albeit slightly slower than littermate-derived control MEFs (Fig. 2c, d, Supplementary Fig. 2 and data not shown). Given our previous results showing that a combined deficiency in *E2f1*, *E2f2*, *E2f3a* and *E2f3b* completely abrogates proliferation of MEFs[8], these results suggest a critical and redundant role for the two *E2f3* isoforms in the control of proliferation. It would thus appear that expression of either *E2f3* isoform is sufficient to support cell proliferation as well as embryonic development.

Next we evaluated the role of E2F activators in postnatal development. Mice deficient for each *E2f3* isoform appeared externally normal, were fertile and lived a normal lifespan (Fig. 1d, Supplementary Fig. 3 and data not shown). We did note, however, that older $E2f3a^{-/-}$ mice had less white adipose tissue (WAT) deposition than aged-matched $E2f3b^{-/-}$ or wild-type mice, lending to their thinner appearance (Fig. 1f and data not shown). Given the observed functional redundancy among E2Fs during embryonic development[8,11,14], we reasoned that loss of additional E2F activators might accentuate the rather mild age-dependent phenotype in $E2f3a^{-/-}$ mice. Both $E2f1^{-/-}$ and $E2f2^{-/-}$ mice were previously shown to have a relatively normal lifespan, but with age these mice developed haematopoietic-related complications[15–19]. As shown in Fig. 2a, $E2f2^{-/-}E2f3a^{-/-}$ offspring were born at the expected frequency, had normal birth weights and matured without any obvious additional defects. Newborn $E2f1^{-/-}E2f3a^{-/-}$ pups were also of normal weight and appearance, but by their third week of life the proliferative index in most tissues was significantly reduced (Supplementary Fig. 4b) and mice became severely runted (Fig. 3a–c). The weight of most organs was reduced in double knockout animals, but this decrease was generally proportional to the decrease in their total body weight (Supplementary

Fig. 4a). Gross inspection, however, revealed a total absence of inguinal and subdermal fat (Fig. 3d-e), an observation that was subsequently extended to all WAT deposits in the body. In contrast, brown adipose tissue was relatively unaffected. Analysis of food intake and faecal secretion ruled out a defect in eating behaviour or fat absorption as a cause for the absence of WAT (Supplementary Fig. 4c). Serum triglycerol levels as well as leptin, a major homeostatic regulator of lipid metabolism, were significantly lower in double knockout mice than in control animals (Supplementary Fig. 4d), suggesting a more direct role of E2F3a function in lipid metabolism. Not surprisingly, most *E2f1*$^{-/-}$*E2f3a*$^{-/-}$ pups died within their first month of life, and the few mice that lived longer were severely incapacitated and could be kept alive only by intensive husbandry (Fig. 3b).

Pathological analysis revealed additional organ defects that could account for the growth retardation, general unhealthy disposition and high morbidity of *E2f1*$^{-/-}$*E2f3a*$^{-/-}$ mice. For example, by 21 days of age, the growth plates in long bones of double knockout mice were severely dysplastic with many cells having mega-nuclei that contributed to poor development of trabeculae (Fig. 3e). This was accompanied by a significant decrease in circulating growth hormone and insulin-like growth factor (IGF)-I levels and a relative increase in brain/body weight ratio (Supplementary Fig. 4a, e), providing an interesting parallel to growth hormone deficiency in humans, which causes dwarfism and mainly affects the growth of limbs but not the skull. Moreover, the outer cortex of their adrenal glands lacked a functional zona fasciculata and their lungs were also abnormal, with a decrease in the branching of the epithelium (Fig. 3e). Young mutant males exhibited severe testicular hypoplasia and females had hypoplastic ovaries that lacked a well-developed corpora lutea (Fig. 3e). As a result the few mice that lived past 2 months of age were infertile. We conclude that multi-organ failure resulted in the general deterioration of health and early death of *E2f1*$^{-/-}$*E2f3a*$^{-/-}$ mice. These results illustrate the critical and redundant roles of *E2f1* and *E2f3a* during postnatal development.

In contrast to *E2f1*$^{-/-}$*E2f3a*$^{-/-}$ mice, *E2f1*$^{-/-}$*E2f3b*$^{-/-}$ and *E2f2*$^{-/-}$*E2f3b*$^{-/-}$ progeny developed normally through puberty and adulthood and had a normal lifespan (Fig. 3b and data not shown). Whereas young *E2f1*$^{-/-}$*E2f3b*$^{-/-}$ mice were slightly smaller than wild-type controls (Fig. 3a and 3c), thorough histological examination failed to detect any of the mutant phenotypes observed in *E2f1*$^{-/-}$*E2f3a*$^{-/-}$ animals, including in WAT, bone, adrenal glands, lungs and gonads (data not shown). These observations suggest important differences between *E2f3a* and *E2f3b* during postnatal development. Analysis of *E2f1*$^{-/-}$*E2f2*$^{-/-}$*E2f3a*$^{-/-}$ and *E2f1*$^{-/-}$*E2f2*$^{-/-}$*E2f3b*$^{-/-}$ triple knockout animals also revealed significant differences between the contribution of *E2f3a* and *E2f3b* towards postnatal development. Whereas both sets of triple knockout embryos could be carried to late stages of gestation, *E2f1*$^{-/-}$*E2f2*$^{-/-}$*E2f3a*$^{-/-}$ embryos died perinatally and *E2f1*$^{-/-}$*E2f2*$^{-/-}$*E2f3b*$^{-/-}$ pups lived well into adulthood, albeit their body weight was reduced (Fig. 2a, b and data not shown). These findings demonstrate that mouse development can proceed in the presence of a single E2F activator, *E2f3a*.

The different roles of *E2f3a* and *E2f3b* during postnatal development could reflect differences in the function of their gene products or differences in the control of their expression[20–22]. To differentiate between these possibilities we used homologous recombination in mouse embryonic stem cells to replace the coding sequence of exon 1a with the coding sequence of exon 1b (*E2f3a*^{3bki}; Fig. 4a, top panel). We were careful to avoid the inadvertent perturbation of regulatory regions that govern expression from the *E2f3a* locus by leaving the 5′ untranslated region and the 3′ splicing junction of exon 1a intact (Fig. 4b). This strategy resulted in the synthesis of an *E2f3a*^{3bki} mRNA that was expressed with a similar cell cycle profile as endogenous *E2f3a* and other known E2F targets, without affecting the expression of endogenous *E2f3b* (Fig. 4d). Strikingly, both *E2f3a*$^{3bki/3bki}$ as well as *E2f1*$^{-/-}$*E2f3a*$^{3bki/3bki}$ mice had normal body weight, WAT deposition, bone structure and growth, adrenal gland morphology and function, as well as normal lung and gonad morphology (Fig. 4h, j, Supplementary Fig. 5 and data not shown). We suggest that the differential requirement for *E2f3a* and *E2f3b* during postnatal development is not based on intrinsic differences between E2F3a and E2F3b protein functions, but rather on differences between how their loci dictate their respective expression.

The above results prompted us to take a similar genetic strategy to determine whether a more distantly related family member, *E2f1*, could also substitute for *E2f3a*. Introduction of the murine *E2f1* coding sequence into exon 1a of the *E2f3a* locus (Fig. 4a bottom panel and Fig. 4c) resulted in the cell-cycle-dependent expression of E2F1 protein (Fig. 4e, f) without affecting expression from the nearby *E2f3b* locus (Fig. 4g). Remarkably, *E2f3a*$^{1ki/1ki}$ as well as *E2f1*$^{-/-}$*E2f3a*$^{1ki/1ki}$ mice lacked any of the phenotypes caused by a deficiency in *E2f3a* and lived to old age (Fig. 4h–j, Supplementary Fig. 5 and data not shown). Together, these findings suggest that the specific role of *E2f3a* in postnatal development is largely predicated by regulatory sequences governing its spatiotemporal expression pattern. However, we can not rule out specific functions for individual E2Fs in tissues that are not critical for development and that were not analysed here. These observations provide a molecular basis for the observed specificity among E2F family members during development.

Many reasons have been provided to explain why mammals have an expanded E2F family when nematodes and flies have a single E2F activator and a single repressor. Extensive genetic and biochemical analysis using cell culture systems has demonstrated that different E2F activators interact with specific cofactors and elicit different biological responses[23–25]. These studies inspired the view that the complexity of the mammalian E2F family affords a more complex and presumably better equipped organism to develop, but *in vivo* evidence in support of this hypothesis has been lacking. The *in vivo* analyses presented here provide new perspectives to this old problem. Our current findings suggest that biological processes regulated by E2F during development in nematodes, flies and mammals are more alike than previously anticipated because, like in nematodes and flies, a single E2F activator (*E2f3a*) is sufficient to support development in the mouse. The surprising observation that E2F3a protein function during postnatal development can be substituted equally well by either the structurally related E2F3b or the more distantly related E2F1 proteins highlights the critical nature of *E2f3a* regulatory sequences for postnatal development. It would appear that beyond broadening gene expression patterns, the mammalian genome has gained little developmental currency by the acquisition of additional E2F family members. It remains possible that the evolution of multiple E2Fs in mammals might represent an adaptation that could serve, beyond simply broadening expression patterns, to meet the challenges faced by ageing animals reared in their natural habitat.

METHODS SUMMARY

Construction of *E2f3a*, *E2f3b*, *E2f3a*^{3bki} and *E2f3a*^{1ki} targeting vectors. The triple *loxP* vector system and standard cloning techniques were used to construct all targeting vectors[8] (see Supplementary Information for details).

Generation of *E2f3a*$^{-/-}$, *E2f3b*$^{-/-}$, *E2f3a*$^{3bki/3bki}$ and *E2f3a*$^{1ki/1ki}$ mice. TC1 129Sv/Ev ES cells were electroporated with 50 μg of NotI-linearized targeting vectors. Homologous recombination was selected with G418 and ganciclovir and verified by Southern blotting using 5′ and 3′ external probes. Appropriate ES clones were injected into e3.5 C57BL/6 blastocysts that were subsequently implanted into foster mothers. We bred the resulting chimaeras with NIH black Swiss females and selected agouti offspring for genotyping. Offspring with germline transmission of the *E2f3a* or *E2f3b* targeted alleles were bred with *EIIa-cre* mice to excise exon 1a or exon 1b, respectively, as well as the neomycin cassette.

Genotyping of knockout and knockin mice. Genotyping of offspring was performed by PCR and Southern blot analysis. Southern blots analysis was performed using genomic DNA digested with EcoRI and a ^{32}P-labelled HindIII-BamHI fragment as the hybridization probe. The wild type, *E2f3a* and *E2f3b* knockout, as well as the *E2f3a*^{3bki} and *E2f3a*^{1ki} alleles, yielded fragments of approximately 6 kb, 5.2 kb, 4.4 kb, 7 kb and 9 kb, respectively. Homozygous mice

were obtained by interbreeding heterozygous mice of each genotype. Multiplex PCR for genotyping used three primers for *E2f3a* (primer 1, primer 2, primer 3) and *E2f3b* (primer 4, primer 5, primer 6), and the same two primers for $E2f3a^{3bki}$ and $E2f3a^{1ki}$ (primer 7; primer 8). Primer sequences are listed in Supplementary Fig. 6.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 November 2007; accepted 7 May 2008.
Published online 25 June 2008.

1. Trimarchi, J. M. & Lees, J. A. Sibling rivalry in the E2F family. *Nature Rev. Mol. Cell Biol.* **3,** 11–20 (2002).
2. Attwooll, C., Lazzerini Denchi, E. & Helin, K. The E2F family: specific functions and overlapping interests. *EMBO J.* **23,** 4709–4716 (2004).
3. Ceol, C. J. & Horvitz, H. R. dpl-1 DP and efl-1 E2F act with lin-35 Rb to antagonize Ras signaling in *C. elegans* vulval development. *Mol. Cell* **7,** 461–473 (2001).
4. Dynlacht, B. D., Brook, A., Dembski, M., Yenush, L. & Dyson, N. DNA-binding and trans-activation properties of *Drosophila* E2F and DP proteins. *Proc. Natl Acad. Sci. USA* **91,** 6359–6363 (1994).
5. Sawado, T. *et al.* dE2F2, a novel E2F-family transcription factor in *Drosophila melanogaster. Biochem. Biophys. Res. Commun.* **251,** 409–415 (1998).
6. DeGregori, J. & Johnson, D. G. Distinct and overlapping roles for E2F family members in transcription, proliferation and apoptosis. *Curr. Mol. Med.* **6,** 739–748 (2006).
7. Helin, K. *et al.* A cDNA encoding a pRB-binding protein with properties of the transcription factor E2F. *Cell* **70,** 337–350 (1992).
8. Wu, L. *et al.* The E2F1–3 transcription factors are essential for cellular proliferation. *Nature* **414,** 457–462 (2001).
9. Humbert, P. O. *et al.* E2f3 is critical for normal cellular proliferation. *Genes Dev.* **14,** 690–703 (2000).
10. Leone, G. *et al.* Identification of a novel E2F3 product suggests a mechanism for determining specificity of repression by Rb proteins. *Mol. Cell. Biol.* **20,** 3626–3632 (2000).
11. Li, J. *et al.* Synergistic function of E2F7 and E2F8 is essential for cell survival and embryonic development. *Dev. Cell* **14,** 62–75 (2008).
12. Giangrande, P. H. *et al.* A role for E2F6 in distinguishing G1/S- and G2/M-specific transcription. *Genes Dev.* **18,** 2941–2951 (2004).
13. Cloud, J. E. *et al.* Mutant mouse models reveal the relative roles of E2F1 and E2F3 *in vivo. Mol. Cell. Biol.* **22,** 2663–2672 (2002).
14. Gaubatz, S. *et al.* E2F4 and E2F5 play an essential role in pocket protein-mediated G1 control. *Mol. Cell* **6,** 729–735 (2000).
15. Yamasaki, L. *et al.* Tumor induction and tissue atrophy in mice lacking E2F–1. *Cell* **85,** 537–548 (1996).
16. Murga, M. *et al.* Mutation of E2F2 in mice causes enhanced T lymphocyte proliferation, leading to the development of autoimmunity. *Immunity* **15,** 959–970 (2001).
17. Field, S. J. *et al.* E2F-1 functions in mice to promote apoptosis and suppress proliferation. *Cell* **85,** 549–561 (1996).
18. Zhu, J. W. *et al.* E2F1 and E2F2 determine thresholds for antigen-induced T-cell proliferation and suppress tumorigenesis. *Mol. Cell. Biol.* **21,** 8547–8564 (2001).
19. Li, F. X., Zhu, J. W., Hogan, C. J. & DeGregori, J. Defective gene expression, S phase progression, and maturation during hematopoiesis in E2F1/E2F2 mutant mice. *Mol. Cell. Biol.* **23,** 3607–3622 (2003).
20. Opavsky, R. *et al.* Specific tumor suppressor function for E2F2 in Myc-induced T cell lymphomagenesis. *Proc. Natl Acad. Sci. USA* **104,** 15400–15405 (2007).
21. Dirlam, A., Spike, B. T. & Macleod, K. F. Deregulated E2f-2 underlies cell cycle and maturation defects in retinoblastoma null erythroblasts. *Mol. Cell. Biol.* **27,** 8713–8728 (2007).
22. Parisi, T. *et al.* Selective requirements for E2f3 in the development and tumorigenicity of Rb-deficient chimeric tissues. *Mol. Cell. Biol.* **27,** 2283–2293 (2007).
23. Blais, A. & Dynlacht, B. D. E2F-associated chromatin modifiers and cell cycle control. *Curr. Opin. Cell Biol.* **19,** 658–662 (2007).
24. Giangrande, P. H., Hallstrom, T. C., Tunyaplin, C., Calame, K. & Nevins, J. R. Identification of E-box factor TFE3 as a functional partner for the E2F3 transcription factor. *Mol. Cell. Biol.* **23,** 3707–3720 (2003).
25. Schlisio, S., Halperin, T., Vidal, M. & Nevins, J. R. Interaction of YY1 with E2Fs, mediated by RYBP, provides a mechanism for specificity of E2F function. *EMBO J.* **21,** 5775–5786 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Moffitt and L. Rawahneh for histology expertise. We also thank J. Nevins, C. Bock and A. Otoshi for support in the generation of the *E2f3a*, *E2f3b*, $E2f3a^{3bki}$ and $E2f3a^{1ki}$ mice, and the Mouse Metabolic Phenotyping center at the University of Cincinnati for advice on the analysis of $E2f1^{-/-}E2f3a^{-/-}$ mice. We are grateful to D. Guttridge, M. Ostrowski and M. Simcox for critically reading the manuscript and helpful suggestions. This work was funded by NIH grants to G.L. (R01CA85619, R01HD042619, R01CA121275, R01HD047470, P01CA097189), to L.W. (K01CA102328), DoD awards to A.d.B. (BC0300893) and J.-L.C. (BC061730), and a T32 fellowship (CA106196) to R.O. G.L. is the recipient of the Pew Charitable Trusts Scholar Award and the Leukemia and Lymphoma Society Scholar Award.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.L. (Gustavo.Leone@osumc.edu).

doi:10.1038/nature07066

METHODS

Construction of *E2f3a*, *E2f3b*, *E2f3a^{3bki}* and *E2f3a^{1ki}* targeting vectors. The *E2f3a* knockout vector contained a neo cassette flanked by *loxP* sites, which was inserted into the HaeII restriction site 108 bp upstream of the *E2f3a* starting codon. An additional *loxP* site was inserted into the HindIII site ~300 bp downstream of exon 1a. The two arms of homology used for recombination included a total 3.1-kb HindIII-EcoRI fragment upstream of the neo cassette and a 900-bp fragment downstream of the 3′ *loxP* site. For the *E2f3b* knockout targeting vector, the same *loxP*–flanked neo cassette was inserted ~800 bp upstream of exon 1b and an additional *loxP* site was inserted ~1.8 kb downstream of exon 1b. The two arms of homology used for recombination included a 1-kb HindIII-HindIII fragment upstream of the neo cassette and a 9-kb EcoRI-XhoI fragment downstream of the 3′ *loxP* site. The *E2f3a^{3bki}* and *E2f3a^{1ki}* knockin constructs were identical to the *E2f3a* knockout construct described above with the following exceptions. For the *E2f3a^{3bki}* construct, the exon 1b open reading frame (ORF) was replaced with the exon 1a ORF. For the *E2f3a^{1ki}* construct, the *E2f1* ORF carrying its own termination codon was inserted downstream of the first ATG in exon 1a. All the final targeting vectors were confirmed by direct DNA sequencing.

Fat absorption assay. Fat absorption assays were performed on 1-month-old mice at the mouse metabolic phenotyping centre at the University of Cincinnati as described previously[26].

Blood serum analysis. Mice at 21 days of age were killed and serum was collected from blood by centrifugation at 6,000*g*. The serum was then stored at −80 °C until further analysis. Triglycerides, cholesterol and leptin were analysed by the Mouse Metabolic Phenotyping Center at the University of Cincinnati. Analysis of serum growth hormone and IGF-I was performed at the Vanderbilt hormone assays and analytical core services.

Generation of MEFs and cell culture conditions. Primary MEFs were isolated from e13.5 embryos using standard methods[8]. MEFs were cultured in DMEM with 15% fetal bovine serum (FBS). Proliferation assays were performed by plating MEFs at 1.45×10^5 cells per 60-mm dish. Duplicate plates were counted daily using a Beckton Dickson Coulter Counter and were re-plated 72 h after at the same initial density. Four independent *E2f1$^{-/-}$ E2f2$^{-/-}$ E2f3a$^{-/-}$* MEF lines and two independent *E2f1$^{-/-}$ E2f2$^{-/-}$ E2f3b$^{-/-}$* MEF lines with their control littermate were used in the proliferation assays. Statistical analysis was performed by pooling all experiments.

Protein and RNA analysis. Cells were scraped from culture dishes in chilled PBS, centrifuged, and washed once with ice-cold PBS. Total protein extracts were prepared by incubating cells in RIPA extraction buffer for 30 min on ice. Total protein was then separated by SDS–PAGE and transferred to PVDF membranes. Blots were probed with antibodies specific for E2F3 (Santa Cruz Sc-878) or E2F1 (Santa Cruz Sc-193); anti-α-tubulin (Sigma T6199) was used to determine protein loading. Total RNA was extracted from MEFs using Qiagen RNeasy mini kit. Reverse transcription of 2 μg of total RNA was performed by combining 1 μl of Superscript III reverse transcriptase (Invitrogen), 4 μl of 5× buffer, 0.5 μl of 100 mM oligo dT primer, 0.5 μl of 25 mM dNTPs, 1.0 μl of 0.1 M dithiothreitol, 1.0 μl of RNase Inhibitor (Roche) and water up to a volume of 20 μl. Reactions were incubated at 50 °C for 60 min and then diluted fivefold with 80 μl of water. Real-time RT–PCR was performed using the BioRad iCycler PCR machine. Each PCR reaction contained 0.5 μl of cDNA template and primers at a concentration of 100 nM in a final volume of 25 μl of SYBR green reaction mix (BioRad). Each PCR reaction yielded only the expected amplicon as shown by the melting-temperature profiles of the final products and by gel electrophoresis. Standard curves were generated using cDNA to determine the linear range and PCR efficiency of each primer pair. Reactions were performed in triplicate and relative amounts of cDNA were normalized to GAPDH. Primer sequences are listed in Supplementary Fig. 6.

Histopathology and immunohistochemistry. Tissue samples were collected and fixed in 10% neutral formalin. Five-micrometre-thick sections were cut and then stained for with haematoxylin and eosin by standard protocols. Immunohistochemistry using Ki-67-specific antibodies (BD Pharmingen 550609) was performed on paraffin-embedded sections.

Statistical analysis. For the proliferation assays in MEFs, a generalized linear model (GLM) was used to study the association between the outcome variable and group. Time was used as a categorical variable, and the interaction 'time' by 'group' was also included in the model. A fixed effect 'experiment' was included to take into account the differences (or variability) among experiments.

26. Jandacek, R. J., Heubi, J. E. & Tso, P. A novel, noninvasive method for the measurement of intestinal fat absorption. *Gastroenterology* **127**, 139–144 (2004).

Vol 454|28 August 2008|doi:10.1038/nature07163

LETTERS

Cell-specific ATP7A transport sustains copper-dependent tyrosinase activity in melanosomes

Subba Rao Gangi Setty[1], Danièle Tenza[2,3], Elena V. Sviderskaya[4], Dorothy C. Bennett[4], Graça Raposo[2,3] & Michael S. Marks[1]

Copper is a cofactor for many cellular enzymes and transporters[1]. It can be loaded onto secreted and endomembrane cuproproteins by translocation from the cytosol into membrane-bound organelles by ATP7A or ATP7B transporters, the genes for which are mutated in the copper imbalance syndromes Menkes disease and Wilson disease, respectively[2]. Endomembrane cuproproteins are thought to incorporate copper stably on transit through the *trans*-Golgi network, in which ATP7A accumulates[3] by dynamic cycling through early endocytic compartments[4]. Here we show that the pigment-cell-specific cuproenzyme tyrosinase acquires copper only transiently and inefficiently within the *trans*-Golgi network of mouse melanocytes. To catalyse melanin synthesis, tyrosinase is subsequently reloaded with copper within specialized organelles called melanosomes. Copper is supplied to melanosomes by ATP7A, a cohort of which localizes to melanosomes in a biogenesis of lysosome-related organelles complex-1 (BLOC-1)-dependent manner. These results indicate that cell-type-specific localization of a metal transporter is required to sustain metallation of an endomembrane cuproenzyme, providing a mechanism for exquisite spatial control of metalloenzyme activity. Moreover, because BLOC-1 subunits are mutated in subtypes of the genetic disease Hermansky–Pudlak syndrome, these results also show that defects in copper transporter localization contribute to hypopigmentation, and hence perhaps other systemic defects, in Hermansky–Pudlak syndrome.

Copper is essential for all cells, but is particularly important for vertebrate pigmentation as a cofactor for the enzyme tyrosinase[5]. Tyrosinase is expressed in epidermal melanocytes and ocular pigment cells, and catalyses the initial steps of melanin biosynthesis within melanosomes[6]. Mutations in either of two copper-binding sites in tyrosinase ablate enzymatic activity and result in oculocutaneous albinism[7]. Melanin synthesis is limited to melanosomes, but tyrosinase is thought to acquire copper within the *trans*-Golgi network (TGN) as it traverses the secretory pathway towards melanosomes, because copper-dependent tyrosinase enzyme activity is first detected in the TGN[8] and in nearby clathrin-coated vesicles[9]. Copper is supplied to tyrosinase by the transporter ATP7A[10]; accordingly, ATP7A-deficient mice are severely hypopigmented[11,12].

Hypopigmentation is a feature of Hermansky–Pudlak syndrome (HPS)[13], a multisystem disorder of subcellular organelle formation. HPS and corresponding mouse models result from mutations in any of 15 genes, most of which encode subunits of cytoplasmic complexes that regulate membrane trafficking. Among them, mutations in components of the eight-subunit BLOC-1 cause the most severe hypopigmentation[14] and an almost complete absence of pigment in melanocytes[15]. Several melanosomal proteins, including tyrosinase-related protein-1 (TYRP1), are notably missorted and excluded from melanosome precursors in BLOC-1-deficient (BLOC-1^{-}) cells[15], but tyrosinase is only partially missorted such that a substantial cohort properly localizes to non-pigmented melanosome precursors[15]. We hypothesized that these organelles lack pigment because the tyrosinase within them is inactive, perhaps owing to lack of its essential cofactor, copper. We thus assessed copper transporter localization in melanocytes and its dependence on BLOC-1.

By immunofluorescence microscopy (IFM) analysis of wild-type pigmented mouse melanocyte cell lines, ATP7A was detected in the perinuclear area and in punctate structures characteristic of the TGN and endosomes as in most other cell types, but was also detected in ~98% of pigmented cells within mature melanosomes, apparent as 'doughnuts' filled with dark melanin (Fig. 1a). These ATP7A-containing structures colocalized with pigment granules detected by bright field microscopy (Fig. 1a, inset) and with the melanosome membrane protein TYRP1 (Fig. 1c, inset). ATP7B was not detected in these cells by IFM (data not shown). The signal obtained with the anti-ATP7A antibody was specific, because it was diminished (in an average of 28%–39% of cells) on depletion of endogenous ATP7A after transduction with several different *Atp7a*-specific small interfering RNAs (siRNAs) or with short hairpin RNA (shRNA) expression constructs (Fig. 1d and Supplementary Fig. 1). Moreover, the signal did not reflect nonspecific adherence to melanin, because many of the ATP7A-depleted, antibody-non-reactive cells retained pigment (probably synthesized before ATP7A depletion; Fig. 1d, e and Supplementary Fig. 1). Immunoelectron microscopy (IEM) analysis confirmed that ATP7A localizes not only to Golgi and TGN membranes (see later in manuscript) but also to the limiting and intralumenal membranes of pigmented melanosomes (Fig. 1f, g). This intramelanosomal distribution resembled that of tyrosinase[16]. ATP7A localization to melanosomes was further verified by subcellular fractionation (see later), consistent with a proteomic analysis of melanosome-enriched subcellular fractions of human MNT-1 melanoma cells[17]. Together, these results establish that a cohort of ATP7A in wild-type pigmented melanocytes localizes to melanosomes.

Whereas a cohort of tyrosinase is present in non-pigmented melanosomes in BLOC-1^{-} cells, other melanosome proteins, including TYRP1, are excluded from melanosomes and accumulate in vacuolar early endosomes[15]. Like TYRP1 (but not tyrosinase[15]), ATP7A was excluded from melanosomes in these cells, because it did not colocalize appreciably with the melanosome precursor marker, PMEL17,

[1]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [2]Institut Curie, Centre de Recherche, Paris F-75248, France. [3]Centre National de la Recherche Scientifique, UMR 144, Paris F-75248, France. [4]Centre for Molecular and Metabolic Signalling, Division of Basic Medical Sciences, St George's, University of London, London SW17 0RE, UK.

using IFM (Supplementary Fig. 3C), and was not detected in fibrillar melanosomes using IEM (Fig. 2g and Supplementary Fig. 4a, b). Instead, ATP7A localized predominantly to the perinuclear area and peripheral early endosomes in BLOC-1^{-} melan-mu and melan-rp cells, as shown by extensive colocalization with the mislocalized TYRP1 using IFM (Fig. 2a and Supplementary Fig. 2; in melan-mu, 39 ± 11% of peripheral ATP7A puncta contained TYRP1 and 81 ± 11% of TYRP1 puncta contained ATP7A; mean ± s.e.m.) and with the early endosomal markers EEA1 and syntaxin 13 (also known as syntaxin 12; Fig. 2b, c and Supplementary Fig. 3A, B). IEM analyses confirmed that ATP7A in these cells localized primarily to Golgi membranes and secondarily to tubulovesicular and vacuolar endosomes—which also often contained TYRP1, internalized transferrin (TRF) and tyrosinase (Fig. 2g–i and Supplementary Fig. 4a,

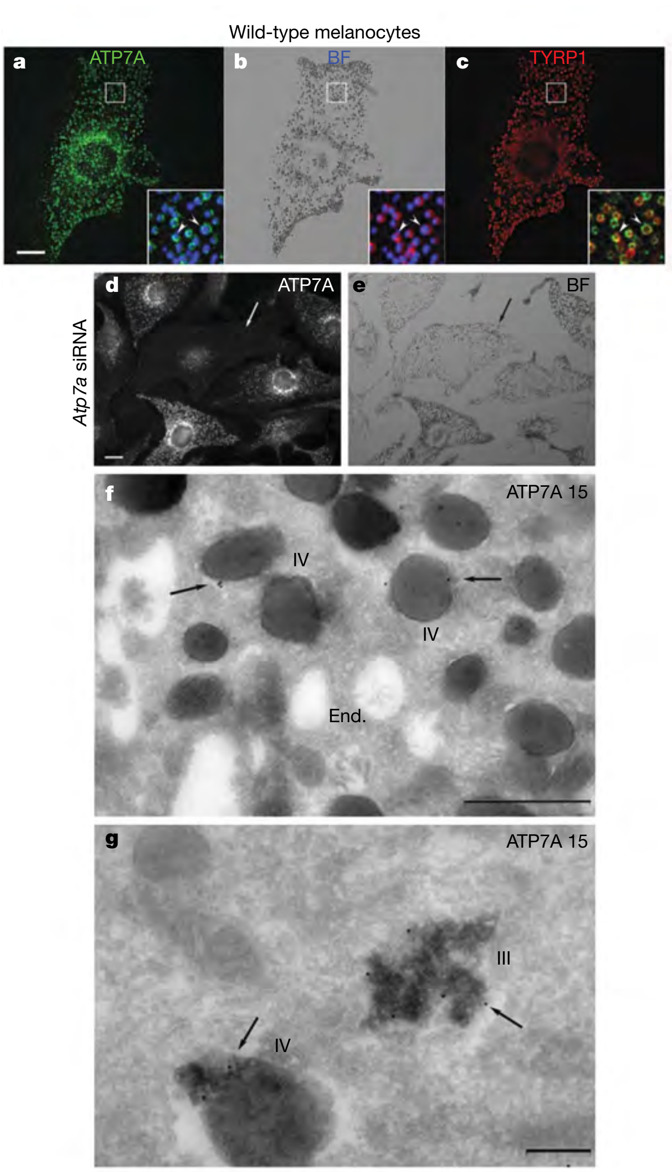


Figure 1 | Copper transporter ATP7A localizes to melanosomes in wild-type melanocytes. **a**–**c**, Fixed melan-a (wild-type) cells were analysed by IFM with antibodies to ATP7A (**a**) and TYRP1 (**c**). The corresponding bright field (BF; **b**) image shows pigmented melanosomes. The insets show ×3-magnified overlays of the boxed regions comparing ATP7A to melanosomes (pseudocoloured blue from an inverted BF image; **a**), melanosomes to TYRP1 (**b**) and ATP7A to TYRP1 (**c**). Arrowheads, colocalized ATP7A, melanosomes and TYRP1. Scale bar, 10 μm. **d**, **e**, Wild-type melan-Ink4a melanocytes were transduced with siRNA to *Atp7a* (*Atp7a-3* siRNA) and analysed 48 h later by IFM for ATP7A and bright field microscopy. Arrow, a cell in which labelling for ATP7A on pigment granules is lost. Scale bar, 10 μm. **f**, **g**, Ultrathin cryosections of wild-type melan-a cells were immunogold labelled with anti-ATP7A antibody (15 nm gold; ATP7A 15) and analysed by electron microscopy. Arrows indicate labelling of stage III and IV melanosomes (III and IV), predominantly on the internal membranes within them. See also Supplementary Fig. 4. End., endosomes; scale bars, 500 nm (**f**) and 200 nm (**g**).

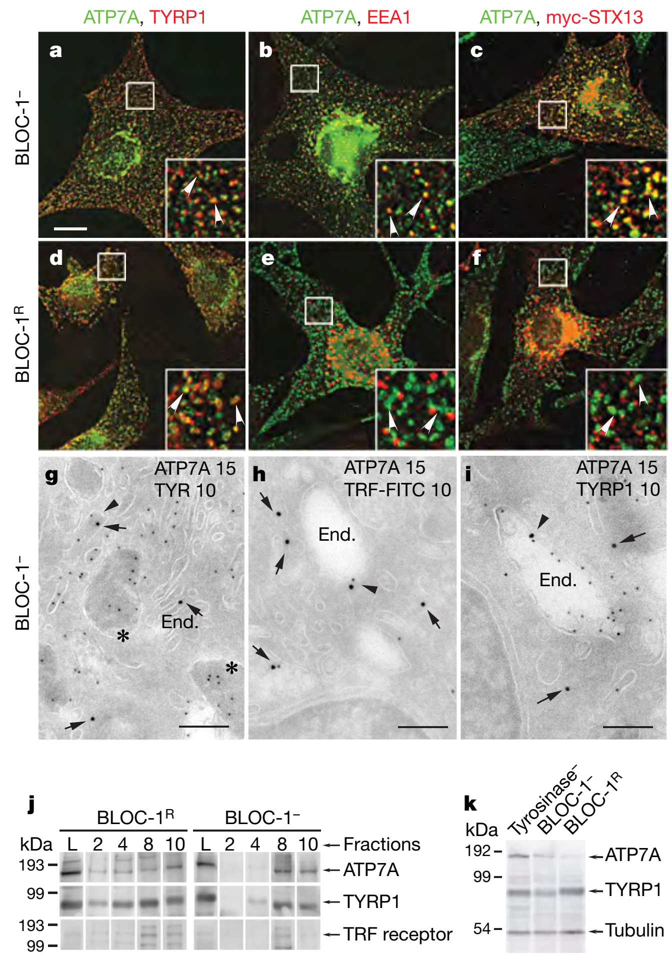


Figure 2 | ATP7A is mislocalized to early endosomes in BLOC-1-deficient melanocytes. **a**–**f**, IFM analysis of BLOC-1^{-} (melan-mu, **a**–**c**) and BLOC-1^{R} (melan-mu:MuHA, **d**–**f**) melanocytes labelled with antibodies to ATP7A and TYRP1 (**a**, **d**), EEA1 (**b**, **e**) or transiently expressed myc epitope-tagged syntaxin 13 (myc-STX13; **c**, **f**). Insets represent boxed regions magnified ×3. Arrowheads, ATP7A colocalized with the indicated marker; scale bar, 10 μm. Note that TYRP1 labels melanosomes in BLOC-1^{R} cells but early endosomes in BLOC-1^{-} cells[15]; EEA1 and myc-STX13 label early endosomes in both cell types. **g**–**i**, Ultrathin cryosections of BLOC-1^{-} melan-mu cells were immunogold labelled with antibodies to ATP7A (15 nm gold; ATP7A 15) and tyrosinase (10 nm gold; TYR 10; **g**), internalized TRF (10 nm gold; TRF-FITC; **h**) or TYRP1 (**i**; 10 nm gold; TYRP1 10) and analysed by electron microscopy. Arrows indicate labelling of ATP7A and arrowheads represent ATP7A with TRF TYR or TYRP1 on endosomes. Asterisks in **g** indicate striated melanosomes labelled for tyrosinase but not ATP7A. End., endosomes; scale bars, 200 nm. See also Supplementary Fig. 4. **j**, Subcellular fractionation of BLOC-1^{-} (melan-mu) and BLOC-1^{R} (melan-mu:MuHA) cells on sucrose step gradients. Eluted fractions (2, 4, 8 and 10), collected from bottom to top, and lysates (L, input loading control) were probed by immunoblot with antibodies to ATP7A, TYRP1 and TRF receptor (to label early endosomes). Note the bands for ATP7A and TYRP1, but not TRF receptor (middle band), in pigment-containing fractions 2 and 4 from melan-mu:MuHA but not melan-mu cells. **k**, Whole cell lysates from tyrosinase-deficient (tyrosinase^{-}, melan-c), BLOC-1^{-} (melan-mu) and BLOC-1^{R} (melan-mu:MuHA) were fractionated by SDS–PAGE and immunoblotted with antibodies to ATP7A, TYRP1 and tubulin.

b)—and to multivesicular late endosomes. BLOC-1 'rescue' by expression of the missing Muted or BLOS3 (also known as BLOC1S3) subunit in melan-mu or melan-rp cells, respectively, resulted in reduced ATP7A expression (Fig. 2k) but nevertheless restored pigmentation and melanosome localization for both ATP7A and TYRP1. IFM showed that 25% ± 7% of peripheral ATP7A localized to TYRP1-containing rings in BLOC-1-rescued (BLOC-1^{R}) melan-mu:MuHA cells, and IEM showed that ATP7A was detected in melanosomes, as well as in endosomes and Golgi (Fig. 2d–f and Supplementary Figs 2 and 4c–e).

IFM and IEM analyses were confirmed by subcellular fractionation on sucrose step gradients, in which a cohort of ATP7A was detected together with TYRP1 in a high density, pigmented melanosome fraction from BLOC-1^{R} cells but not from BLOC-1^{-} cells (Fig. 2j and Supplementary Fig. 4f). In the latter, ATP7A was only detected in low-density fractions and cofractionated with the TRF receptor and mistargeted TYRP1 (Fig. 2j and Supplementary Fig. 4f). Furthermore, the results of cross-linking and immunoprecipitation analyses from both BLOC-1^{-} and BLOC-1^{R} cells show that a cohort of ATP7A associates with the BLOC-1-dependent cargo protein TYRP1, but not BLOC-1-independent cargoes LAMP1 or TRF receptor (Supplementary Fig. 5a–c). Although the interaction with TYRP1 is not required for ATP7A trafficking to melanosomes (Supplementary Fig. 5d), the results suggest that ATP7A and TYRP1 are in close proximity within a transport intermediate before BLOC-1 function. Together, the data establish that melanosomal localization of ATP7A requires BLOC-1. In contrast, adaptor protein 3 (AP-3) is dispensable for ATP7A trafficking to melanosomes, because a large cohort of ATP7A, like TYRP1 (ref. 15), efficiently localized to pigmented melanosomes in AP3B1-deficient melan-pe cells derived from HPS type 2 model pearl mice (Supplementary Fig. 2m–o).

If ATP7A is required to supply copper to tyrosinase in melanosomes, then the cohort of tyrosinase in non-pigmented melanosomes of BLOC-1^{-} cells might be inactive owing to a lack of copper resulting from the absence of ATP7A. Indeed, 3,4-dihydroxy-phenylalanine (DOPA) cytochemistry analyses showed that tyrosinase activity in BLOC-1^{-} melan-mu cells was nearly as low as that in tyrosinase-deficient melan-c cells, as judged by the paucity of melanin deposits formed in the presence of the tyrosinase substrate L-DOPA relative to the negative control D-DOPA (Fig. 3a, b, g, h). In contrast, L-DOPA-induced melanin deposits were clearly observed in BLOC-1^{R} melan-mu:MuHA cells (Fig. 3c, d) and AP3B1-deficient melan-pe cells (Fig. 3i, j). The low tyrosinase activity in BLOC-1^{-} cells did not result from tyrosinase instability because immunoblotting showed that tyrosinase protein levels were only modestly lower in BLOC-1^{-} than in BLOC-1^{R} cells, and much higher than the residual unprocessed tyrosinase in melan-c cells (Fig. 3k). Modest tyrosinase activity was detected in melan-rp cells (Fig. 3e, f), which lack the BLOC-1 subunit BLOS3; this is consistent with the partial assembly of BLOC-1 in these cells[18] and the milder hypopigmentation of reduced pigmentation mice from which melan-rp cells are derived[18,19]. Together, these results indicate that tyrosinase is largely inactive in BLOC-1^{-} melanocytes.

To determine whether tyrosinase in BLOC-1^{-} melanocytes is inactive because it lacks copper, we tested whether the addition of copper to the post-fixation DOPA cytochemistry reaction restored activity. Indeed, L-DOPA-dependent melanin pigmentation was notably increased in BLOC-1^{-} melan-mu cells by the addition of 20 μM copper sulphate (Fig. 4a–d). Pigmentation of BLOC-1^{R} melan-mu:MuHA cells also increased on copper addition, albeit not as notably (Supplementary Fig. 6a–d). To determine the subcellular location of tyrosinase activity in these cells, DOPA cytochemistry reactions were analysed by electron microscopy. As expected, L-DOPA-induced melanin deposits (indicating active tyrosinase) in BLOC-1^{R} melanocytes were observed in mature (stage III and IV) melanosomes (Fig. 4g, h) and in the TGN in both the absence and the presence of copper, although copper consistently increased the signal in the TGN (Fig. 4g, h); a cohort (~26%–28%; Supplementary Fig. 7g) of melanosomes

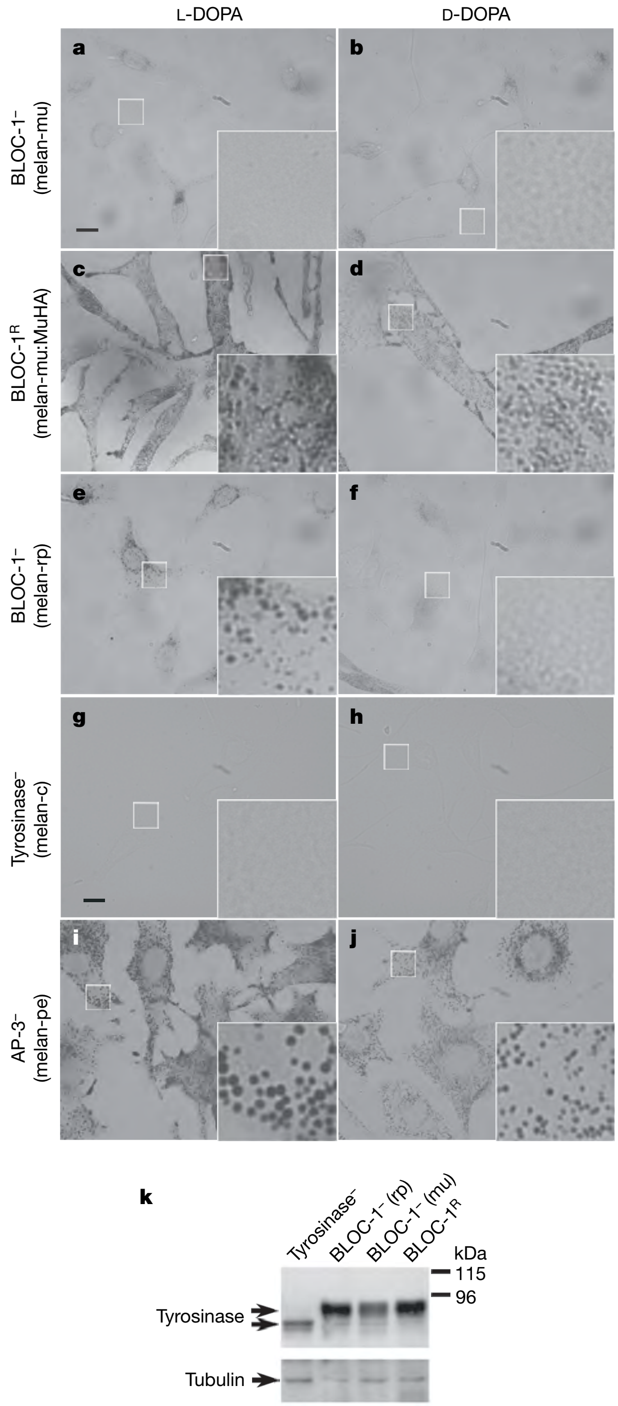


Figure 3 | Tyrosinase is present but inactive in BLOC-1-deficient melanocytes. a–j, Bright field microscopy analysis of BLOC-1^{-} (melan-mu, **a**, **b**; melan-rp, **e**, **f**), BLOC-1^{R} (melan-mu:MuHA, **c**, **d**), AP-3-deficient (melan-pe, **i**, **j**) and tyrosinase mutant (melan-c, **g**, **h**) mouse melanocytes treated with L- or D-DOPA for 3 h. Tyrosinase activity is indicated by melanin deposition in the presence of L-DOPA but not D-DOPA. Insets show boxed regions magnified ×5; scale bars, 10 μm. Note melanin deposits were not observed in melan-mu (**a**) or control melan-c (**g**) cells, and modest deposits were observed in melan-rp cells in which BLOC-1 is partially assembled (**e**). **k**, Whole cell lysates of mouse melanocytes were fractionated by SDS–PAGE and immunoblotted with anti-tyrosinase antibody. Blots were reprobed with anti-tubulin antibody as a loading control. Note that tyrosinase protein levels are only mildly reduced in BLOC-1^{-} melan-mu cells relative to BLOC-1^{R} melan-mu:MuHA cells.

remained unpigmented (Supplementary Fig. 7a, b), consistent with the absence of tyrosinase from stage II melanosomes[16]. In contrast, without added copper, L-DOPA-induced melanin deposits were only observed in the TGN of BLOC-1^{-} melan-mu cells and not in the striated melanosomes (Fig. 4e and Supplementary Fig. 7c, e, g). Addition of copper to the L-DOPA cytochemistry reaction resulted in further melanin deposition within most striated melanosomes in BLOC-1^{-} cells (Fig. 4f and Supplementary Fig. 7d, f); the fraction of melanosomes containing pigment among these cells was comparable to that of wild-type cells (Supplementary Fig. 7g). The pigmented organelles observed in the presence of copper probably correspond to slightly more electron-dense fibrillar organelles observed in the absence of copper (Supplementary Fig. 7c) and that harbour both PMEL17 and tyrosinase[15]. Together, these data indicate that tyrosinase is initially activated in the TGN, where ATP7A accumulates in these cells (Supplementary Fig. 4a–e), but is inactive in the nonpigmented melanosomes specifically because of a lack of copper. Copper also stimulated tyrosinase activity in BLOC-1^{-} cells within tubular and vacuolar endosomes (Supplementary Fig. 7d, f), to which a cohort of tyrosinase is mislocalized in these cells[15]. Despite the presence of 25%–50% of immunodetectable tyrosinase in early endosomes of wild-type melanocytes[15,16], very little enzyme activity was detected in endosomes of BLOC-1^{R} cells regardless of the addition of copper (Supplementary Fig. 7a, b). Because the TGN and tubular endosomes are intermediates in tyrosinase transport to melanosomes[16], the data suggest that tyrosinase inefficiently binds copper ions in the TGN *in vivo* and loses its bound copper ions in endosomes *en route* to melanosomes.

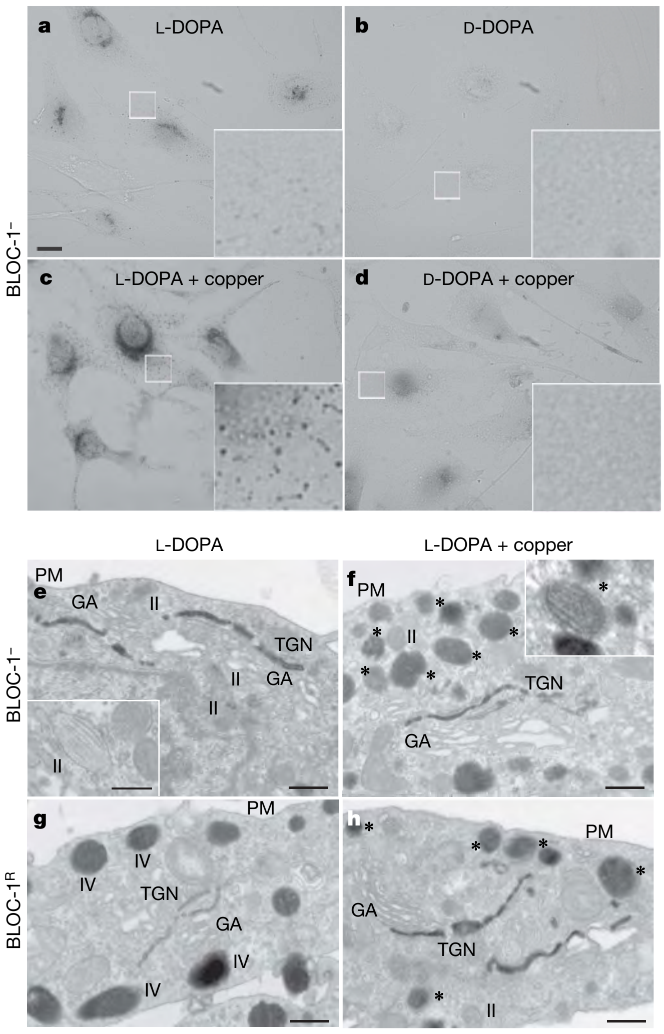


Figure 4 | Copper restores *in vitro* tyrosinase activity in melanosomes of BLOC-1-deficient melanocytes. a–h, DOPA cytochemistry of BLOC-1^{-} melan-mu (**a**–**f**) or BLOC-1^{R} melan-mu:MuHA (**g**, **h**) cells in the absence or presence (+ copper) of 20 μM copper sulphate in the reaction buffer. **a**–**d**, Bright field microscopy analysis of cells treated for 4 h as indicated. Insets, boxed regions magnified ×5; scale bars, 10 μm. Note the increased melanin deposits in BLOC-1^{-} cells in the presence (**c**) compared to the absence (**a**) of copper. **e**–**h**, Electron microscopy analysis of thin sections of cells treated for 2 h. Note melanin deposits in the *trans*-most Golgi cisternae, but not in striated melanosomes, of BLOC-1^{-} melan-mu cells in the absence of excess copper (**e**; inset), and the extra deposition of melanin in striated melanosomes in the presence of copper (**f**, asterisks; inset). II, IV, melanosome stages II and IV; GA, golgi apparatus; PM, plasma membrane; TGN, *trans*-Golgi network. Scale bars, 500 nm and inset, 200 nm.

Our studies show that tissue-specific localization of the copper transporter ATP7A to a lysosome-related organelle, the melanosome, is required to supply copper to sustain the activity of a resident metalloenzyme, tyrosinase. We confirm that tyrosinase is first loaded with copper in the TGN[8], a normal trafficking intermediate *en route* to melanosomes. However, unlike for ceruloplasmin, which binds copper very stably[20], the copper seems to be loaded onto tyrosinase inefficiently and is subsequently stripped—probably in endosomal intermediates that function in TGN-to-melanosome trafficking[16]—and must be reloaded within mature melanosomes. This mechanism provides tight spatial control of tyrosinase activity to ensure that melanin is produced only in mature melanosomes and that endosomal transport intermediates and melanosome precursors are protected from toxic melanin intermediates generated by premature tyrosinase activity (the TGN might additionally lack the L-DOPA or tyrosine substrates). Similar spatial regulation of copper integration probably exists for other endomembrane metalloproteins such as peptidylglycine α-amidating monooxygenase[21], and preliminary results (our unpublished data) suggest that copper release from tyrosinase can be regulated by extracellular cues to provide fine-tuning of cuproenzyme activity. How bound copper is destabilized within endosomes is not clear, but a contributing factor is probably the low pH of these compartments. Tyrosinase is inactive at acidic pH[22], perhaps owing to protonation of copper-coordinating histidine residues within the copper-binding domain[23], and even the addition of excess copper *in vitro* could not restore tyrosinase activity at a pH of 5 in BLOC-1^{-} melanocytes (Supplementary Fig. 8). However, prolonged exposure to low pH does not seem to be sufficient to destabilize bound copper (data not shown), suggesting that extra factors participate. In contrast, melanosomes become more alkaline as they mature[24], providing an environment conducive to the reactivation of tyrosinase on re-exposure to copper.

The localization of ATP7A to melanosomes requires BLOC-1, which facilitates tissue-specific transport of selected cargoes to melanosomes[15] and other lysosome-related organelles[13,25–27]. This dependence in part explains the hypopigmentation of BLOC-1-deficient mice and human patients with HPS types 7 and 8. Mislocalization of metal transporters such as ATP7A or ATP7B might contribute to other cell type defects in HPS, including the loss of functional dense granules in platelets and of lamellar bodies in type II pneumocytes. Retinal pigment epithelial cells, which also harbour melanosomes and are hypopigmented in HPS, express both ATP7A and ATP7B[28]. Whether either or both transporters localize to melanosomes during the brief period of melanogenesis in these cells[29] remains to be tested. Notably, tyrosinase activity in BLOC-1^{-} melanocytes could be restored by addition of copper *in vitro* (Supplementary Fig. 7i), but not by incubation of live cells in excess copper (data not shown). This suggests that alternative transporters on melanosomes either cannot substitute for ATP7A or are also mislocalized in BLOC-1^{-} cells.

METHODS SUMMARY

For bright field and immunofluorescence microscopy, cells were fixed with 2%–4% formaldehyde, labelled with the indicated primary and Alexa Fluor-conjugated secondary antibodies and analysed on a DM IRBE microscope (Leica

Microsystems) equipped with an Orca digital camera (Hamamatsu) using OpenLab software (Improvision). Images obtained from consecutive *z*-planes were processed using subtractive volume deconvoluton with Improvision OpenLab. DOPA cytochemistry was performed as described[30] and similarly analysed but without image deconvolution; all images were obtained at similar camera and illumination settings. Electron microscopy analyses of immunogold-labelled ultrathin cryosections or of DOPA-treated cells, and immunoblotting of whole cell lysates, were performed as described[18,19]. Melan-Ink4a cells were transfected with siRNAs using oligofectamine and either analysed two to three days later (Fig. 1) or transfected a second time on day 3 and analysed two days later (Supplementary Fig. 1). Subcellular fractionation of cells disrupted by Dounce homogenization was performed using sedimentation on sucrose step gradients.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 January; accepted 10 June 2008.
Published online 23 July 2008.

1. Thiele, D. J. Integrating trace element metabolism from the cell to the whole organism. *J. Nutr.* **133,** 1579S–1580S (2003).
2. Mercer, J. F. The molecular basis of copper-transport diseases. *Trends Mol. Med.* **7,** 64–69 (2001).
3. Yamaguchi, Y. *et al.* Biochemical characterization and intracellular localization of the Menkes disease protein. *Proc. Natl Acad. Sci. USA* **93,** 14030–14035 (1996).
4. Harris, E. D. Cellular copper transport and metabolism. *Annu. Rev. Nutr.* **20,** 291–310 (2000).
5. Lutsenko, S. *et al.* Function and regulation of human copper-transporting ATPases. *Physiol. Rev.* **87,** 1011–1046 (2007).
6. Marks, M. S. & Seabra, M. C. The melanosome: membrane dynamics in black and white. *Nature Rev. Mol. Cell Biol.* **2,** 738–748 (2001).
7. Oetting, W. S. *et al.* Oculocutaneous albinism type 1: the last 100 years. *Pigment Cell Res.* **16,** 307–311 (2003).
8. Novikoff, A. B., Albala, A. & Biempica, L. Ultrastructural and cytochemical observations on B-16 and Harding–Passey mouse melanomas. The origin of premelanosomes and compound melanosomes. *J. Histochem. Cytochem.* **16,** 299–319 (1968).
9. Maul, G. G. & Brumbaugh, J. A. On the possible function of coated vesicles in melanogenesis of the regenerating fowl feather. *J. Cell Biol.* **48,** 41–48 (1971).
10. Petris, M. J., Strausak, D. & Mercer, J. F. The Menkes copper transporter is required for the activation of tyrosinase. *Hum. Mol. Genet.* **9,** 2845–2851 (2000).
11. Levinson, B. *et al.* The mottled gene is the mouse homologue of the Menkes disease gene. *Nature Genet.* **6,** 369–373 (1994).
12. Mercer, J. F. *et al.* Mutations in the murine homologue of the Menkes gene in dappled and blotchy mice. *Nature Genet.* **6,** 374–378 (1994).
13. Wei, M. L. Hermansky–Pudlak syndrome: a disease of protein trafficking and organelle function. *Pigment Cell Res.* **19,** 19–42 (2006).
14. Gautam, R. *et al.* Interaction of Hermansky–Pudlak syndrome genes in the regulation of lysosome-related organelles. *Traffic* **7,** 779–792 (2006).
15. Setty, S. R. *et al.* BLOC-1 is required for cargo-specific sorting from vacuolar early endosomes toward lysosome-related organelles. *Mol. Biol. Cell* **18,** 768–780 (2007).
16. Theos, A. C. *et al.* Functions of adaptor protein (AP)-3 and AP-1 in tyrosinase sorting from endosomes to melanosomes. *Mol. Biol. Cell* **16,** 5356–5372 (2005).
17. Chi, A. *et al.* Proteomic and bioinformatic characterization of the biogenesis and function of melanosomes. *J. Proteome Res.* **5,** 3135–3144 (2006).
18. Starcevic, M. & Dell'Angelica, E. C. Identification of snapin and three novel proteins (BLOS1, BLOS2, and BLOS3/reduced pigmentation) as subunits of biogenesis of lysosome-related organelles complex-1 (BLOC-1). *J. Biol. Chem.* **279,** 28393–28401 (2004).
19. Gwynn, B. *et al.* Reduced pigmentation (*rp*), a mouse model of Hermansky–Pudlak syndrome, encodes a novel component of the BLOC-1 complex. *Blood* **104,** 3181–3189 (2004).
20. Sato, M. & Gitlin, J. D. Mechanisms of copper incorporation during the biosynthesis of human ceruloplasmin. *J. Biol. Chem.* **266,** 5128–5134 (1991).
21. De, M. *et al.* Trafficking of a secretory granule membrane protein is sensitive to copper. *J. Biol. Chem.* **282,** 23362–23371 (2007).
22. Wang, N. & Hebert, D. N. Tyrosinase maturation through the mammalian secretory pathway: bringing color to life. *Pigment Cell Res.* **19,** 3–18 (2006).
23. Martinez, J. H. *et al.* The involvement of histidine at the active site of Harding–Passey mouse melanoma tyrosinase. *Biochem. Int.* **11,** 729–738 (1985).
24. Raposo, G. *et al.* Distinct protein sorting and localization to premelanosomes, melanosomes, and lysosomes in pigmented melanocytic cells. *J. Cell Biol.* **152,** 809–824 (2001).
25. Raposo, G. & Marks, M. S. Melanosomes—dark organelles enlighten endosomal membrane transport. *Nature Rev. Mol. Cell Biol.* **8,** 786–797 (2007).
26. Raposo, G., Marks, M. S. & Cutler, D. F. Lysosome-related organelles: driving post-Golgi compartments into specialisation. *Curr. Opin. Cell Biol.* **19,** 394–401 (2007).
27. Salazar, G. *et al.* BLOC-1 complex deficiency alters the targeting of adaptor protein complex-3 cargoes. *Mol. Biol. Cell* **17,** 4014–4026 (2006).
28. Krajacic, P. *et al.* Retinal localization and copper-dependent relocalization of the Wilson and Menkes disease proteins. *Invest. Ophthalmol. Vis. Sci.* **47,** 3129–3134 (2006).
29. Lopes, V. S. *et al.* Melanosome maturation defect in Rab38-deficient retinal pigment epithelium results in instability of immature melanosomes during transient melanogenesis. *Mol. Biol. Cell* **18,** 3914–3927 (2007).
30. Boissy, R. E., Zhao, Y. & Gahl, W. A. Altered protein localization in melanocytes from Hermansky–Pudlak syndrome: support for the role of the HPS gene product in intracellular trafficking. *Lab. Invest.* **78,** 1037–1048 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Gitlin, M. Petris, B. Eipper, S. Lutsenko and A. Peden for reagents; A. Dancis, L. King and C. Burd for comments; and D. Harper for technical assistance. This work was supported by National Institute of Health grants R01 EY015625 and R21 GM078474 (to M.S.M.), CNRS, Institut Curie and Fondation pour la Recherche Médicale (to G.R.), Wellcome Trust program grant 064583 (to E.V.S. and D.C.B.), and postdoctoral fellowship 0625437U from the American Heart Association (to S.R.G.S.).

Author Contributions S.R.G.S. designed and performed most of the experiments, prepared most of the figures, wrote an initial draft of the manuscript and participated in all stages of manuscript revision. D.T. performed all of the electron microscopy analyses and provided valuable insights into data interpretation. E.V.S. and D.C.B. provided the cell lines used in all of the experiments, performed confirmatory experiments and guided others, provided further data not shown in the manuscript and participated in manuscript revision. G.R. oversaw the electron microscopy analyses, prepared electron micrographs for the figures, contributed substantially to experimental design and participated in manuscript revision. M.S.M. oversaw the entire project, designed many of the experiments in collaboration with S.R.G.S., coordinated work among collaborators and participated in all stages of manuscript revision.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.S.M. (marksm@mail.med.upenn.edu).

METHODS

Reagents. Chemicals were from Sigma-Aldrich except where noted. Tissue culture reagents and Oligofectamine were from Invitrogen. FuGENE 6, Hygromycin B and protease inhibitor cocktail were from Roche Diagnostics. Protein A conjugated to 10- and 15-nm gold particles (PAG10 and PAG15) was from Cell Microscopy Center.

Antibodies. Monoclonal antibodies: TA99 to TYRP1 and 9E10 to myc were from American Type Culture Collection; 1D4B to mouse LAMP-1 was from Developmental Studies Hybridoma Bank; HMB-45 to PMEL17 was from Lab Vision; H68.4 to TRF receptor was from Zymed; and GTU88 to γ-tubulin was from Sigma-Aldrich. Polyclonal rabbit antibodies: H-90 to TYRP1 (used for immunoblotting) was from Santa Cruz Biotechnology; anti-fluorescein isothiocyanate (FITC) was from Invitrogen; and anti-tyrosinase[18] and anti-ATP7A (used for IFM studies)[31] have been described. Goat anti-EEA1 was from Santa Cruz Biotechnology. To generate Menkes antibody, Menkes peptide, corresponding to the carboxy-terminal 21 amino acid of mouse ATP7A (accession number NM 009726; sequence, SEPDKHSLLVGDFREDDDTTL), was synthesized and used to produce polyclonal antisera in rabbits (Genemed Synthesis). Crude antisera were used for immunoblotting experiments. For IEM experiments, antisera were affinity purified using Menkes peptide coupled to SulfoLink coupling gel (Pierce) according to the manufacturer's instructions.

DNA constructs and siRNAs. Human syntaxin 13 with an N-terminal myc-epitope tag was amplified from I.M.A.G.E. clone 3851266 and subcloned into the BamHI and XhoI sites of pCDNA3.1(+) (Invitrogen). *Atp7a* shRNA constructs were generated by cloning oligodeoxyribonucleotide duplexes containing the target sequences into the BamHI–HindIII sites of pRS shRNA (OriGene Technologies). The following sequences were selected as targets: AAA TGGCATGACTTGTAATTC (*Atp7a-1*), AATCTATGGGTTTAGAAGTAG (*Atp7a-2*) and AAAGATCGATCAGCCAACCAC (*Atp7a-3*). All plasmid inserts were verified by DNA sequencing. siRNAs bearing identical target sequences were obtained from Qiagen Sciences, Inc. As a control, empty pRS shRNA plasmid or siRNA to human t-Golgin-1 (ref. 32) was used in shRNA or siRNA experiments, respectively.

Cell culture, transgene expression and siRNA transduction. The following immortalized mouse melanocyte cell lines were used. Wild-type melan-a[33] and melan-Ink4a[32] from C57BL/6 mice, tyrosinase-deficient melan-c and TYRP1-deficient melan-b[33], AP-3-deficient melan-pe[16], BLOC-1-deficient melan-mu3 (melan-mu) and melan-rp2 (melan-rp) and rescued melan-mu:MuHA and melan-rp:BLOS3 (stably expressing their respective missing BLOC-1 subunit)[15], have been described. BLOC-1-deficient (BLOC-1^{-}) and BLOC-1 rescued (BLOC-1^{R}) cells in the text refer to melan-mu and melan-mu:MuHA, respectively, unless otherwise stated. All cells were maintained as described[32] except without added cholera toxin. For shRNA knockdown, melan-Ink4a cells were transiently transfected using FuGene-6 with 2 μg of each shRNA construct separately in a 6 well dish and analysed 48 h post-transfection. For siRNA knockdown, melan-Ink4a cells were transiently transfected twice using oligofectamine as described[34] and analysed one to two days after the second transfection. For transient expression of myc-syntaxin 13, melan-mu and melan-mu:MuHA cells were transfected using FuGene-6 with 2 μg of plasmid in a 6 well dish. Cells were analysed by IFM 48 h later; analysis was limited to cells expressing low levels of the transgene.

***In vitro* tyrosinase activity.** DOPA cytochemistry was carried out as described previously[30]. In brief, for light microscopy analyses, cells on coverslips were freshly fixed with 4% formaldehyde, washed thrice with PBS and incubated at 37 °C for 2–4 h in PBS containing 0.1% L-DOPA or D-DOPA (as a negative control) with or without 20 μM copper sulphate. The assay was carried out in either PBS pH 7.4 or citrate phosphate buffer pH 6.0 or pH 5.0, with or without 20 μM copper sulphate. Cells were washed with PBS and analysed by bright field microscopy as described using identical camera settings. For electron microscopy, cells were fixed for 1–2 h with Karnovsky's fixative (72 mM sodium cacodylate pH 7.4, 4% paraformaldehyde and 4 mM calcium chloride) containing 0.5% glutaraldehyde and rinsed 2–3 times with cacodylate buffer (0.2 M sodium cacodylate, pH 7.4, 0.1 M calcium chloride). Cells were incubated at 37 °C for 2 h in cacodylate buffer containing 0.1% L-DOPA or D-DOPA with or without 20 μM copper sulphate, and then washed 2–3 times and fixed overnight with Karnovsky's fixative containing 2% glutaraldehyde. Cells were rinsed, incubated for 1 h in cacodylate buffer containing 1% osmium tetroxide and 1.5% potassium ferrocyanide, washed, dehydrated and embedded for sectioning as described[35].

Immunofluorescence microscopy and immunoelectron microscopy. IFM and IEM studies were carried out as described[15]. To quantify marker overlap using IFM, deconvolved paired images were rendered binary by density slicing and the total area of overlap between them was calculated for objects containing more than 5 pixels, excluding the densely labelled perinuclear area. Eleven to eighteen cell profiles comprising >5,000 objects were quantified for each pairwise comparison. For IEM, TRF-FITC was internalized for 15 min and detected with anti-FITC antibody. Quantification of the fraction of pigmented melanosomes was determined by counting 228–303 oblong compartments each in epon-embedded sections of melan-mu and melan-mu:MuHA cells, treated with L-DOPA with and without added copper.

Immunoblotting and subcellular fractionation. For immunoblotting of whole cell lysates, cells were lysed at 4×10^7 cells ml^{-1} in 100 mM Tris-HCl pH 7.6, 100 mM NaCl, 1% SDS and protease inhibitor cocktail, sonicated for 1–2 min in a water bath sonicator, clarified by centrifugation, fractionated and immunoblotted as described[36]. For subcellular fractionation, cells were washed with PBS, suspended in 0.25 M sucrose buffer (0.25 M sucrose, 1 mM EDTA, 25 mM HEPES, pH 7.4, 0.02% sodium azide and protease inhibitor cocktail) and then homogenized using a hand-held Dounce homogenizer. Cell homogenates were clarified by centrifugation at 700*g* at 4 °C for 10 min and fractionated on a sucrose step gradient (2.0 M, 1.6 M, 1.4 M, 1.2 M and 1.0 M sucrose densities from bottom to top) at 160,000*g* at 4 °C for 4 h in a Beckman L-80 ultracentrifuge using a SW41Ti rotor. Fractions were collected from bottom to top using a peristaltic pump. Aliquots of each fraction were further fractionated by SDS–PAGE and probed by immunoblotting with antibodies to ATP7A, TYRP1 and TRF receptor. The melanosome fraction in wild-type cells, assessed by pigmentation and quantified by A_{600}, appears largely at the interfaces between 2.0 M and 1.4 M sucrose densities.

Co-immunoprecipitation. Cells were collected, washed once with PBS, suspended in bicine buffer (0.1 M bicine pH 8.0, 50 mM sodium chloride, 1 mM magnesium chloride and protease inhibitor cocktail) with or without dithiobis(-succinimidylpropionate) (Pierce) at indicated concentrations, and then incubated on ice for 30 min. Cells were then lysed in 1% Triton X-100 or 6% CHAPS on ice for 10 min. Cross-linking reactions were neutralized by the addition of Tris-HCl pH 7.4 (25 mM) at room temperature for 10 min and then clarified by centrifugation at 20,800*g* at 4 °C for 20 min. Lysates were pre-incubated for 30 min at 4 °C with protein A or G-Sepharose (GE Healthcare Bio-Sciences) pre-bound to normal rabbit sera, and then incubated for 10–16 h at 4 °C with protein A or G-Sepharose pre-bound to anti-TYRP1 antibody or to anti-TRF receptor or anti-LAMP1 antibodies as negative controls. After the incubation, beads were washed twice with wash buffer (0.1 M Tris-HCl, pH 7.4, 50 mM sodium chloride, 1 mM magnesium chloride and either 0.1% Triton X-100 or 0.6% CHAPS) and suspended in SDS–PAGE sample buffer. Cross-linked products were cleaved with dithiothreitol (30 mM) and incubated at 50 °C for 20 min. All samples were fractionated by SDS–PAGE and immunoblotted with antibodies to ATP7A, TYRP1 and either TRF receptor or LAMP-1.

31. Hamza, I., Prohaska, J. & Gitlin, J. D. Essential role for Atox1 in the copper-mediated intracellular trafficking of the Menkes ATPase. *Proc. Natl Acad. Sci. USA* **100**, 1215–1220 (2003).
32. Sviderskaya, E. V. *et al.* p16(Ink4a) in melanocyte senescence and differentiation. *J. Natl. Cancer Inst.* **94**, 446–454 (2002).
33. Bennett, D. C. *et al.* Cloned mouse melanocyte lines carrying the germline mutations albino and brown: complementation in culture. *Development* **105**, 379–385 (1989).
34. Yoshino, A. *et al.* tGolgin-1 (p230, golgin-245) modulates Shiga-toxin transport to the Golgi and Golgi motility towards the microtubule-organizing centre. *J. Cell Sci.* **118**, 2279–2293 (2005).
35. Reynolds, E. S. The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. *J. Cell Biol.* **17**, 208–212 (1963).
36. Berson, J. F. *et al.* Pmel17 initiates premelanosome morphogenesis within multivesicular bodies. *Mol. Biol. Cell* **12**, 3451–3464 (2001).

naturejobs

The course of true love never did run smooth — and the same can often be said of a relationship with science. At last week's meeting of the American Chemical Society in Philadelphia, Pennsylvania, a careers session heard one such story of highs and lows.

Rachel Pytel discussed two personal timelines. Her career timeline was fairly typical — graduate school, recruitment on-campus, careers fairs — but her "attitude timeline" was a lot more revealing. When she began graduate school in 2002, Pytel summed up her attitude with the phrase: "Yay, science!" By 2003, disillusionment had set in and it was: "Man, I stink at this. Everyone's smarter than me." By 2004, she'd decided she didn't hate science, but she hated doing research. But Pytel would again refine her assessment, and she later decided that doing research was fine, as long it was applied and team-oriented. She elected to avoid academia and government and took a job at materials-science company Rohm and Haas.

At a separate session, other chemists discussed how they found career paths that both fuel their love of science and make them excited to go to work every day. Derrick Hamlin uses his master's degree in chemistry to bolster his law practice. It has helped him defend clients in drug, drink-driving and environmental cases. In one case, he exonerated a client facing arson charges by telling the jury about the evaporation rate of chemicals left behind when burning gasoline rather than kerosene.

Jennifer Mass uses her chemistry training in art conservation — investigating the chemistry behind art degradation and finding ways to restore damaged works. She and her colleagues, for example, showed that one reason for patches of faded colour in Matisse's painting *The Joy of Life* is that he used paints containing cadmium rather than the more expensive chromium.

For Pytel, the key was not finding an 'unconventional' job but planning her career search. She attributes her satisfaction with her career to patiently attending careers fairs, preparing for job interviews and carefully reflecting on what pace she wanted and skills she possessed. Only then did she once again exclaim: "Yay, science!"

Gene Russo is editor of *Naturejobs*.

CONTACTS
Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW: Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:** Tom Hancock
***Naturejobs* online production:** Dennis Chu

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

India
Vikas Chawla (+91 1242881057)
e-mail: v.chawla@nature.com

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe (+81 3 3267 8765)
e-mail: a.watanabe@natureasia.com
Business Development Manager, Greater China/Singapore:
Gloria To (+852 2811 7191)
e-mail: g.to@natureasia.com

Spamface

Down in the jungle, something stirs.

Martin Hayes

It was the first big story of my career. I'd graduated in 2022 and, after a couple of months bumming around, I'd managed to land a job at gomergomergomer.com. The pay was lousy and so were the hours but the editor hardly ever interfered with my articles. I was 25, stone broke and living in my pal Warren's junk room, but I was also incredibly happy to be an actual, real-life working journalist.

JACEY

I got the story because no one else wanted it — five weeks in the Brazilian rain forest with the Lost Earth Project. The LEP had secured a hefty grant to track down and study the half-dozen or so still-uncontacted tribes living in the forest. It had been decided in 2008 that all efforts would be made to prevent these tribes from coming into contact with the outside world. This obviously ruled out study and observation, but in the intervening years technology had advanced to the point at which the LEP could incorporate declassified Military Invisibility Shield technology into a light and packable material. The plan was to use this new technology, in the form of invisible tents and hides, to monitor a tribe near the border with Peru. The LEP people would film them day and night, document their activities and customs, and I would be there, documenting the film-makers, taking notes for what was planned to be a four-issue series on ggg.com.

It had taken more than a year to get authorization for the expedition. People were understandably worried that there might be some catastrophic clash of cultures. The LEP team consisted of three men, two women and their mascot, an AI raccoon called Spamface. He was one of the early series-theta models, not indistinguishable from a real raccoon but close enough and pretty sharp. They used him as a kind of personable computer: he recorded their conversations, translated the local dialect, transmitted their research, made the coffee and cracked dirty jokes. He walked on his hind legs and wore a small waist-pack in which he stored his cell-phone, a cigarette lighter (although I never saw him smoke) and a random collection of old electrical components. He was treated like any other team member.

The invisible hides worked well and we set up a small filming post about four hundred yards from the village. We'd only been there three days when things began to go wrong.

Spamface started acting a little screwy, he'd forget to send the previous day's footage, make basic errors in his calculations, and the jokes just kept getting dirtier and dirtier. Steve, the team leader, thought that maybe the high humidity levels were playing havoc with his circuits.

On the fifth day we woke as usual at 6 a.m.. Spamface was gone, and coming from the forest was a kind of low, monotonous chant.

Steve fired up the camera and zoomed in on the commotion.

Spamface was sitting nonchalantly on the steps of the chief-elder's hut, completely surrounded by chanting tribesmen. The women were all kneeling and rocking back and forth. Spamface crossed his legs and began telling one of his jokes and soon they were all crying with laughter.

Steve's first instinct was to rush into the camp, grab the pesky raccoon, poke his finger into its right ear and flick the off switch. He couldn't though: contact with the tribe was forbidden. He'd have to get permission from both his bosses and the Brazilian government and without Spamface's high-speed transmitter that could take hours, even days.

By sundown Spamface had shown them how to make alcohol; used a complicated system of pulleys, tracks and differently sized pieces of fruit to explain the heliocentric theory; built a small but functional dynamo from the odds and ends in his waist-pack; and started showing them old episodes of *Mr Bean* on his phone. They loved it.

Within 24 hours they had constructed an elaborate throne and it soon became clear that Spamface was being worshipped as a god. He had his pick of the women, although, thankfully, he was a first-generation model and so lacked the equipment to carry the offer through. Some of the men began tattooing Spamface's chuckling visage onto their chests.

They really did believe him to be a god. And who could blame them — a remote tribe who were met with a small, ferociously intelligent raccoon who just happened to speak the local dialect, who could create light in the darkness simply by winding the handle of a small machine, who could show them moving images of a man in a tweed jacket gurning and driving badly, and who imparted all his knowledge to them freely, without any ulterior motive.

They were spending up to seven hours a day kneeling and chanting at the foot of his throne when we finally received permission to go in under cover of darkness and get him the hell out of there.

We retrieved him without too much trouble, Steve was bitten on the wrist while trying to deactivate him but that was as bad as it got. By sun-up on the following day, we were ten miles away.

The whole thing had been such a fiasco that ggg.com decided to spike the article — eight hard days in the jungle, and all for nothing.

I kept track of the story though, and when news of the lost tribe broke, some missionaries tracked them down and tried to convert them. The tribe threw them out of the village. The chief was later reported to have said that: "A talking racoon was one thing, but outlandish stories of loaves, fishes and virgin births was simply too much mumbo-jumbo to be believed." ■

Martin Hayes lives in Arklow, a town on the east coast of Ireland. His short stories have appeared in various magazines and anthologies. He also writes comics. You can visit him at www.paroneiria.com.

FUTURES